THE WORLD BANK ECONOMIC REVIEW

Volume 6 January 1992 Number [illegible]

THE WORLD BANK ECONOMIC REVIEW

The World Bank Economic Review is a professional journal for the dissemination of World Bank-sponsored research that informs policy analyses and choices. It is issued three times a year, with an occasional additional issue devoted to a special area of current research or policy interest.

The *Review* is directed to an international readership among economists and social scientists in government, business, and international agencies, as well as in universities and development research institutions. The *Review* emphasizes policy relevance and operational aspects of economics, rather than primarily theoretical and methodological issues. It is intended for readers familiar with economic theory and analysis but not necessarily proficient in advanced mathematical or econometric techniques. Articles will illustrate how professional research can shed light on policy choices. Inconsistency with Bank policy will not be grounds for rejection of an article.

Articles will be drawn primarily from work conducted by World Bank staff and consultants. Before being accepted for publication by the Editorial Board, all articles are reviewed by two referees who are not members of the Bank's staff; articles must also be recommended by at least one external member of the Editorial Board.

The *Review* may on occasion publish articles on specified topics by non-Bank contributors. Any reader interested in preparing such an article is invited to submit a proposal of not more than two pages in length to the Editor.

The views and interpretations in articles published are those of the authors and do not necessarily represent the views and policies of the World Bank or of its executive directors or the countries they represent. When maps are used, the designations employed are solely for the convenience of the reader and do not imply the expression of any opinion whatsoever on the part of the World Bank or its affiliates concerning the legal status of any country, territory, city, or area, or concerning the delimitations of its boundaries or national affiliation.

Comments or brief notes responding to *Review* articles are welcome and will be considered for publication to the extent that space permits. Please direct all editorial correspondence to the Editor, *The World Bank Economic Review*, The World Bank, Washington, D.C. 20433, U.S.A.

Subscriptions to *The World Bank Economic Review* are available without charge to readers with mailing addresses in developing countries and in socialist economies in transition. Written request is required every three years to renew such subscriptions. See the coupon at the back of this issue for subscription charges and additional information.

Manufactured in the United States of America
ISSN 0258-6770

This journal is indexed regularly in *Current Contents/Social & Behavioral Sciences*, *Index to International Statistics*, *Journal of Economic Literature*, *Public Affairs Information Service*, and *Social Sciences Citation Index*®.

THE WORLD BANK ECONOMIC REVIEW

Volume 6 January 1992 Number 1

SYMPOSIUM ON INFLATION IN SOCIALIST ECONOMIES IN TRANSITION

ARTICLES

THE WORLD BANK ECONOMIC REVIEW

Volume 6 May 1992 Number 2

THE WORLD BANK ECONOMIC REVIEW

Volume 6 September 1992 Number 3

A Symposium on Inflation in Socialist Economies in Transition

This symposium draws primarily on papers presented at the Conference on Managing Inflation in Socialist Economies, sponsored by the Economic Development Institute of the World Bank, in Laxenburg, Austria, March 1990. The articles in the symposium were refereed in the usual way. The Editorial Board invited the organizer of the conference, Simon Commander, to write the introduction to the symposium.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 3-12

Inflation and the Transition to a Market Economy: An Overview

Simon Commander

European socialist economies were commonly characterized by low or negligible open inflation, full employment, and stable relative prices and real incomes. Associated features were significant repressed or hidden inflation and disequilibria in goods markets. More recently, as economic and political reform has advanced, attention has shifted to translating repressed into open inflation. Where such reforms have proceeded against a background of a monetary overhang, the implications of that overhang for demand-side policies have figured prominently. In all instances, however, the key underlying issue has been the transmission mechanism for inflation once the initial impulse associated with price liberalization has been imparted.

Although it is widely recognized that major price liberalization has to be central to any stabilization, there is considerable debate concerning the optimal pace and the dynamic effects of price reforms. The arguments for liberalization turn on the importance of achieving a rational, relative price structure that can eliminate imbalances in the goods market while enabling allocative efficiency gains. Yet, as recent experience in Eastern Europe testifies, a variety of price liberalization schemes have involved significant jumps in the price level that translate into acceleration in the rate of increase of prices. In addition, developments on the real side have been seriously adverse, with large declines in recorded output. There is further evidence of an initially slow, but accelerating, increase in the rate of open unemployment.

The articles in this symposium concentrate on the problem of inflation in socialist economies in transition. Given recent experience in several reforming socialist economies with high inflation, Dornbusch draws out the main comparative lessons from other high-inflationary experiences, particularly in Latin America. Stabilizing inflation requires fiscal correction and reforms to the tax system. There is commonly a role for incomes policy in achieving a more rapid and coordinated disinflation, but only as a complement to, rather than a substitute for, budget balancing.

Simon Commander is with the Economic Development Institute of the World Bank. He would like to thank Sara Calvo and Peter Knight for helpful comments on an earlier draft.

Commander and Coricelli use data from Hungary and Poland to analyze the propagation mechanisms for inflation in a partially reformed economy. Developments on the cost side are critical in relating exogenous, policy-determined adjustments to the price level to increases in the rate of inflation. Further, the absence of conventional equilibrating mechanisms and of market-based restraints on prices and wages gives rise to underlying inflationary pressure, indicating, among other things, the importance of incomes policy in any stabilization.

Calvo deals with the role of interest rate policy in a stabilization program and outlines the possible adverse consequences of applying high interest rates over protracted periods. If a stabilization program does not enjoy full credibility, the contraction that high rates are supposed to induce may not result. This is likely to have particular relevance for the reforming socialist economies, given segmentation in the financial system and the absence of mechanisms to direct household savings to the firm sector.

This theme is further developed by Calvo and Coricelli, who invoke interest rates and credit policy as major explanatory factors behind both output decline and inflation persistence in the Polish stabilization. Technological and credit market shocks, rather than demand-side developments, are given prominence, which raises questions about the consequences of standard macroeconomic policy prescriptions in a reforming socialist economy.

This introduction emphasizes the initial conditions of the prereform socialist economies and the mechanisms by which inflationary pressures are carried over into the accelerated reform period. Recent experiences with price reform and stabilization are summarized, with attention focused on the phenomenon of persistent inflation.

I. The Prereform Socialist Economies

In a centrally planned economy decisions are largely based on considerations of quantity. The partially reformed economy accords a greater role to (constrained) market-based rules, with progressive price liberalization and the emergence of dual pricing systems. Bulgaria, Czechoslovakia, and Romania in 1989–90 correspond to the former system; Hungary and Poland in the 1980s to the partially reformed case. In both regimes unemployment was not tolerated, monetary policy was basically passive, and enterprises not only did not have to behave as profit maximizers but, to a certain degree, could be indifferent to financial performance.

Inflationary tensions in these systems had different manifestations. It was widely held that endemic shortages provoked excess demand in both goods and factor markets due to fixed or sticky prices and wages and that this resulted in repressed inflation. Rationing occurred for both households and enterprises. The gap between notional and effective demand at official prices in goods markets translated into excess liquid balances or excessive inventories (Nuti 1986).

Labor shortages resulting from imposed full employment caused enterprises to hoard labor. Along with the institutional bargaining relationships and the relatively weak bargaining strength of managers, the earlier systems were marked by powerful wage pressure on costs. Enterprises, particularly in partially reformed economies, pursued explicit cost-plus pricing routines. The pace at which cost increases were passed through to consumers depended on how automatic product-specific markups were and hence, in part, on the degree of the planner's or government's aversion to price increases.

In this setting, and with high levels of concentration in industry without administrative controls over price and wage changes, inflation (open or repressed) could accelerate rapidly. In short, no endogenous equilibrating mechanism existed in the system. The underlying inflationary mechanism could thus be traced back to a combination of factors. The low productive efficiency of enterprises and the regional structures of production and trade associated with the Council for Mutual Economic Assistance (CMEA) arrangement generated shortages and imbalances in goods markets and hence chronic sellers' markets. Further, the soft budget constraints of the enterprise sector and the banking system fueled increases in subsidy payments and provoked fiscal crises. With pervasive price controls, exogenous price shocks tended to widen the gap between domestic and international prices, thus creating more distorted relative price structures and fiscal pressures.

Table 1 presents data on inflation in the European socialist economies and in industrial and developing countries during 1971–91. It is evident that selective price liberalization during the 1980s was associated with rising open inflation. The response of both centrally planned and partially reformed economies to the external shocks of the 1980s was generally to depress investment, maintain broadly stable real wages, and raise subsidies to enterprises through the budget, negative interest rates, or preferential credit allocations. The deteriorating performance of the enterprise sector and associated growth in subsidies had to be covered by inflationary finance. Servicing of the large external debt, restricted access to foreign borrowing in the 1980s, and limited instruments for placing domestic debt increased the rate of money creation associated with a given noninterest deficit.

High inflationary episodes have occurred in Poland and Yugoslavia and, more recently, in the Soviet Union. In all cases the causal link between inflation and budget deficits is strong when money financing is the only option. Moreover, as Dornbusch emphasizes, not only will there be profound nonlinearity in the link between inflation and the deficit, but the real fiscal deficit is not exogenous because inflation determines real revenues. Inflation not only erodes revenues (Tanzi-Olivera effect), but in reforming socialist economies widespread withholding of taxes by enterprises—the bedrock of the tax system—further deepens the fiscal crisis. An accommodating monetary stance traces back to the dynamics of subsidy policy resulting from price controls, the gap between domestic and external prices, and the pervasive soft budgets of the enterprises. Piecemeal price liberalization has tended to be strongly inflationary when fiscal correction was

Table 1. *Inflation in Eastern Europe, Industrial Countries, and Developing Countries, 1971–90*
(percent)

Country or country group	*Average, 1971–80*	*1981*	*1982*	*1983*	*1984*	*1985*	*1986*	*1987*	*1988*	*1989*	*1990*	*1991*
Bulgaria	2.0	0.5	0.3	1.5	0.7	1.8	3.9	0.1	1.2	6.2	64.0	330.0[a]
Czechoslovakia	1.1	0.9	4.7	1.1	0.9	1.3	0.4	0.1	0.2	1.4	10.0	52.3[b]
Romania	0.9	2.0	16.9	5.2	1.1	0.4	2.0	0.4	3.0	0.6	4.7	80.0[a]
Soviet Union	0.3	1.0	3.0	1.0	−1.0	0.7	2.0	1.3	0.6	2.0	5.3	100.0[a,d]
Hungary	4.5	4.6	6.9	7.3	8.7	7.0	5.3	8.2	16.3	17.0	28.9	26.6[b]
Poland	4.6	24.4	101.5	23.0	15.7	14.4	18.0	25.2	60.2	245.3	249.3	34.2[b]
Yugoslavia	20.0	39.8	31.5	40.2	54.7	72.4	90.0	221.0	294.0	1,340.0	130.0	49.4[a]
Industrial countries	8.6	10.1	7.6	5.2	4.7	4.1	2.4	3.0	3.4	4.5	5.0	2.0[a]
Developing countries	20.6	28.6	30.1	39.9	35.1	35.5	27.0	35.4	57.3	75.9	102.7	12.4[c]

a. First 6 months.
b. January–July 1991.
c. January–April 1991.
d. IMF estimate.
Source: World Bank data.

insufficient to offset subsidies generated by the gap between controlled and market prices.

In Poland and Yugoslavia high inflation was preceded by flight from domestic money that would have increased the income velocity of domestic M1 and eroded the inflation tax base, thus raising equilibrium inflation (Coricelli and Rocha 1991). The accelerated dollarization characteristic of these cases reflected basic uncertainty regarding the direction of economic policy and the apparent lack of rules in determining price changes by the government. Anticipation of further price changes resulted in upward adjustments to market prices, thus widening the gap between administered and market prices, adversely acting on the fiscal deficit, and resulting in overshooting.

In Yugoslavia currency depreciation, triggered by closing access to external credits to finance current account deficits and de facto wage indexation, proved a powerful inflationary channel. The inflationary spiral was triggered, and in part driven, by nominal exchange rate and wage increases consistent with real exchange rate and real wage targets and with the inflation tax generated by the foreign exchange target. However, although the central government accounts show no major fiscal imbalances, inclusion of the quasi-fiscal deficit yields a more conventional picture. Enterprise claims on the banking system expanded significantly through interest rate subsidies (Mates 1990; Rocha 1991). Real exchange rate shocks further resulted in enterprises being unable to service their foreign debts. Consequently, socialization of external liabilities and credit subsidies fueled the expansion of the quasi-fiscal deficit and was a major factor behind monetary expansion.

De facto wage indexing raised the responsiveness of inflation to price shocks. A normal cost output pricing rule and a real wage target provoking a ratchet effect likely account for raising the inflation rate. An accommodating monetary policy would strengthen this result. Indeed, for Poland it appears that even in the high inflation period of 1989, wage stickiness, price controls, and an effectively indexed exchange rate served to maintain inflation inertia. Further, with indexation a real depreciation would immediately feed through pricing and the budget into accelerated inflation.

Piecemeal economic reforms that weakened centralized controls on both prices and wages tended to expose the latent inflationary pressures associated with the ownership and control structure of the economy. The weak competitive environment associated with high levels of concentration and market power and the underlying tendency for wage drift derived from either an explicit worker management system—as in Yugoslavia—or de facto worker bargaining power—as in Poland—are important explanatory factors. The coexistence of market and controlled prices with the associated emergence of nonsynchronization in price determination further imparted inertia. To the extent that the government adjustment rule for controlled prices was targeted on excess demand, the likely outcome through expectations was to destabilize the system (Commander and Coricelli 1990).

II. Stabilization since 1990

Since January 1990 most of the socialist economies in transition have attempted to put in place a stabilization program, a central component of which has been generalized price revisions leading to large upward shifts in the price level. Such programs have also explicitly aimed at a large contraction in subsidies and consequent fiscal correction alongside tight monetary policy and, in some instances, establishment of an independent central bank to signal the repudiation of the earlier accommodating monetary policy.

In Czechoslovakia and Poland the exchange rate has been fixed to provide a nominal anchor to the stabilization. An additional nominal wage anchor has also been employed, with wage cuts enforced ex ante to avert a price-wage spiral. In Bulgaria the exchange rate has been allowed to float, with wages and money providing the main anchors. In Romania money has been used as the main anchor, with wages providing a secondary nominal anchor. A dual exchange rate system was established in 1990, with convergence to a unified rate in 1991. Current account or internal convertibility has been established—except in Bulgaria and Romania—with varying degrees of trade opening.

The stabilizations have aimed at very significant contractions in net domestic credit. In all instances output has fallen sharply. The collapse of the Soviet market and of CMEA trading relations appears to have accounted for some of the decline in output, but this is only a partial explanatory factor with a different incidence at both country and sectoral levels. A major decline in output preceded the dismantling of the CMEA and the oil price shock of 1990 in almost all cases.

Wages have tended to fall sharply at the start of the stabilization, followed by some recovery. Unemployment has risen from a negligible base to a range of 6–12 percent and is projected to accelerate more rapidly. In general, output has declined more sharply than unemployment has risen, resulting in reduced labor productivity. Relatively little privatization has been achieved as yet, and there has been minor formal bankruptcy of firms. The structure of the economies has remained broadly intact, but enterprises have in principle been forced to adapt their pricing and other behavior in response to the exogenous shift in rules.

Persistent Inflation

Large jumps in the price level, generally followed by periodic further revisions to remaining administered goods' prices, have generally eliminated any monetary overhang and shortages. In Poland a comprehensive trade opening resulted in the fairly immediate import of a new relative price structure (Lipton and Sachs 1990). But in Hungary and Poland price level effects have been followed by inflationary persistence. In the Polish case, 18 months after the start of the stabilization, the price level was nearly five times higher, output and real wages were roughly 40 percent lower, unemployment was at around 10 percent, but inflation was still running at 3–4 percent a month.

The persistence of inflation can be related to the combined role of expectations and credibility. Anticipation of further price jumps for incompletely liber-

alized goods, such as energy and transportation, could be expected to feed back into current pricing behavior. Similarly, the structural stability of the system, given little privatization, and lags in effectively importing competition could continue to promote inflationary pressures characteristic of the partially reformed economies. This could be expected to show up in the pricing behavior of enterprises and in the evolution of the markup, as well as in the behavior of wages. Strong wage pressure and slower than warranted employment adjustments would point to a lack of underlying credibility in the program as workers test the resolve of managers and government to enforce a hard budget constraint on firms. It can be reasonably assumed that any tradeoff between wage and price changes and unemployment is only partially or perversely perceived by agents.

Calvo and Coricelli argue that the sources of persistent output contraction and inflation can be traced to the enterprise sector and rigidities in the financial system. Two initial conditions are particularly important. First, the financial system is segmented; there are different circuits for enterprises and households and no ready mechanism for channeling household savings to the enterprise sector. Second, there is a stock of interenterprise debt, reflecting one facet of the soft budget constraint. The principal impulse is through a supply-side shock, which raises input prices very significantly. A monetary crunch through tight credit ceilings and high interest rates has a fairly immediate impact on firms by raising the cost of working capital. This disqualifies firms from producing at full capacity and causes output contraction, labor redundancies, and the expansion of the interfirm credit market (see also Hrncir and Klacek 1991; Tardos 1990). The ex ante stock of interenterprise debt acts as a recessionary transmission mechanism, aside from complicating the discriminating role of a monetary squeeze across "good" and "bad" firms (Calvo and Frenkel 1991).

Technological shocks through input price rises and credit market shocks that impart a liquidity squeeze may in part explain the depth and persistence of output reduction but does less well in explaining the persistence of inflation. On the cost side, it appears that the average markup over costs (net of subsidies) declined during 1990 but that this decline was concentrated in the last quarter (Schaffer 1991). Because wages fell significantly in the earlier part of the year, it seems likely that an undervalued fixed exchange rate allowed greater space for inflationary pricing and corresponding appreciation of the exchange rate. Price rather than wage shocks have dominated. There have been significant lags in the adjustment of domestic prices to international prices alongside nontrivial domestic relative price realignments (Blanchard and Layard 1991).

Wage Pressure and Incomes Policy

Available evidence indicates that autonomous wage-push or price-push through the raising of profit margins have not been major explanatory factors behind stubbornly high inflation rates in several of these economies during the most recent period. There is evidence, however, of reemerging wage-push pressures, particularly in Poland and Romania.

The potential for wage drift is of concern especially because of uncertainty

over ownership and future profitability, given new sets of relative prices. To the extent that tight money and relatively high interest rates are associated with lower output, in turn yielding lower employment, probabilities of unemployment and loss of access to labor or product market rents will feed back into current decisions by workers. This can be expected to have clear repercussions for wage claims, especially given the weight of labor-dominated firms in these economies. In such firms worker control can be treated as tantamount to a powerful trade union presence, where wages and employment are subject to a joint maximization. Workers will not be constrained directly by the labor demand curve but by the firm's profit level. Wages will tend to equate average rather than marginal product but will also depend on the degree of government aversion to bail-outs and hence the soft budget constraint (Commander, Coricelli, and Staehr 1991).

The institutional core and the absence of true performance-related discipline could be expected to fuel wage demands, particularly as workers seek to test the resolve of any reforming government to adhere to its announced policy goals of fiscal balance and tight money. A reforming government starts with low credibility; it will need to shock workers and managers into behaving like profit maximizers and may thus welcome selective bankruptcies or partial closures to demonstrate commitment. Associated losses in output and employment would likewise be accepted.

The institutional legacy and the tendency for wage drift emanating from decentralization and uncertainty provide powerful arguments for formal methods of wage regulation. There is no doubt that wage regulation will imply efficiency costs and impede the appropriate adjustment of relative wages, but the macroeconomic costs of abandoning wage controls in an environment in which the rules of the game are not fully perceived will also not be trivial. Consequently, a range of tax-based incomes policies have been applied. In Poland tax penalties were levied on above-norm wage payments, in which the object of regulation in 1990 was the firm's wage bill. Similarly, in Bulgaria, Czechoslovakia, Hungary, and Romania tax-based incomes policies have been implemented (Commander and Staehr 1991). An explicit incomes policy seems essential for attaining macroeconomic targets in the absence of a more complete introduction and comprehension of standard market-based equilibrating mechanisms. More difficult is the question of how long such controls should be maintained and the impact of an emerging private sector on the relative wage structure, labor allocation, and aggregate employment (Blanchard 1991).

Institutional arrangements are crucial in determining the efficiency—both short- and longer-run—with which inflationary wage claims are contained. In this realm, some clear cross-country differences have emerged, with different, implicit longer-run models of wage bargaining as their underpinnings. In Poland there appears to be the expectation that a centrally imposed incomes policy would give way to a competitive labor market endogenously generating wage behavior consistent with a low-inflation regime. In Bulgaria and Czechoslovakia, by contrast, more emphasis has been given to creating a tripartite struc-

ture of wage bargaining, which encompasses trade unions, managers, and government. In Czechoslovakia the outcome of the 1991 bargain was to agree upon a level of real wage reduction broadly consistent with the macroeconomic targets while also restricting the range of uncertainty over real wage movements given major price shocks. The latter will be an important feature when information regarding the size and timing of shocks is incomplete, which will guard in part against high nominal wage claims fueled by anticipation of future price rises. The basic model resembles more the corporatist structure of the Nordic countries and Austria. As elsewhere, the ability to hold to agreed norms would depend on the ability to sustain a basic consensus among the partners, the willingness of individual unions to relinquish bargaining powers to a central union body, and the extent of unanticipated price shocks.

III. Concluding Remarks

Controlling inflation remains at the heart of the reforms in the erstwhile socialist economies. The problem is more than usually difficult given the need for price liberalization and a new relative price structure. The microeconomic features of the economies and the legacy of the past—as reflected in ownership arrangements, wage setting, organization of the financial system, and allocative mechanisms—risk nullifying or distorting the effects of conventional macroeconomic policies. Credit policy appears to be an important case in hand. At the same time conventional associations, such as a Phillips curve, remain largely absent. In a context of uncertainty over ownership and employment status, unemployment cannot necessarily be expected to exercise restraint on wage demands; instead it may be associated with accelerated wage claims, subject only to the restraint placed by incomes policy.

Further, rapid growth in unemployment and associated output losses have led to correspondingly sharp rises in claims on the budget by firms and the unemployed, thus raising doubts about commitment to fiscal correction. Supply shocks have been significant in all reforming economies and have been associated with a decline in labor productivity. Sharp adjustments to relative prices, as through a major increase in input prices, and higher nominal interest rates could lower output and promote a decline in labor productivity. The scale of decline would depend on the willingness of enterprises to shed labor, this willingness itself being a function of the degree of competitive pressure faced by the enterprises. Large undervaluation of the exchange rate at the outset of the program would provide an initial cushion and impart some measure of price rigidity, particularly in sectors—such as capital goods—in which competition was restricted.

As the experience of Poland and other countries testifies, prices cannot instantaneously jump to equilibrium values. Even with trade opening, domestic prices have appeared to adjust rather slowly to international prices, while staggered price adjustments of sensitive goods—such as housing and transportation—require further periodic shifts in the price level. The combination of lags in

relative price adjustments and powerful oil price and subsequent CMEA shocks appear to be important factors in accounting for the persistence of inflation.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Blanchard, O. J. 1991. "Notes on the Speed of Transition, Unemployment, and Growth in Poland." Massachusetts Institute of Technology, Department of Economics, Cambridge, Mass. Processed.

Blanchard, O. J., and Richard Layard. 1991. "Post-Stabilization Inflation in Poland." Massachusetts Institute of Technology, Department of Economics, Cambridge, Mass. Processed.

Calvo, Guillermo A., and J. A. Frenkel. 1991. "From Centrally Planned to Market Economy: The Road from CPE to PCPE." *IMF Staff Papers* 38 (2, June): 268–99.

Commander, Simon, and Fabrizio Coricelli. 1990. "The Macroeconomics of Price Reform in Socialist Countries: A Dynamic Framework." PRE Working Paper 555. World Bank, Economic Development Institute, Washington, D.C. Processed.

Commander, Simon, Fabrizio Coricelli, and Karsten Staehr. 1991. "Wages and Unemployment in the Transition to a Market Economy." In Georg Winckler, ed., *Central and Eastern Europe: Roads to Growth*. Washington, D.C.: International Monetary Fund.

Commander, Simon, and Karsten Staehr. 1991. "The Determination of Wages in Socialist Economies: Some Microfoundations." PRE Working Paper 713. World Bank, Economic Development Institute, Washington, D.C. Processed.

Coricelli, Fabrizio, and Roberto Rocha. 1991. "Stabilization Programs in Eastern Europe: A Comparative Analysis of the Polish and Yugoslav Programs of 1990." PRE Working Paper 732. World Bank, Country Economics Department, Washington, D.C. Processed.

Hrncir, Miroslav, and Jan Klacek. 1991. "Stabilisation Policies and Currency Convertibility in Czechoslovakia." *European Economy* Special Edition No. 2:17–39.

Lipton, David, and Jeffrey Sachs. 1990 "Creating a Market Economy in Eastern Europe: The Case of Poland." *Brookings Papers on Economic Activity* 1: 75–147.

Mates, Nevin. 1990. "Inflation in Yugoslavia: Specific Forms of Public Deficit Caused by Parafiscal Operations of the Central Bank." Economics Institute, Zagreb. Processed.

Nuti, D. M. 1986. "Hidden and Repressed Inflation in Soviet-Type Economies: Definitions, Measurements and Stabilization." *Contributions to Political Economy* 5: 37–82.

Rocha, Roberto. 1991. "Inflation and Stabilization in Socialist Countries: Some Lessons from the Yugoslav Experience." In Simon Commander, ed., *Managing Inflation in Socialist Economies in Transition*. EDI Seminar Series. Washington, D.C.: World Bank.

Schaffer, Mark. 1991. "A Note on the Polish State-Owned Enterprise Sector in 1990." Centre for Economic Performance, London School of Economics. Processed.

Tardos, Martin. 1990. "The Present State of the Transforming Hungarian Economy." Paper prepared for Vienna Institute for Comparative Economic Studies Workshop on East-West European Economic Integration. Processed.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 13-31

Lessons from Experiences with High Inflation

Rudiger Dornbusch

In economies where price control has been the rule, the most serious concern may be recognition of the inflation problem. Beyond the initial correction of subsidies there is the broader issue of the risk of a serious inflation. This article looks at the problem of high inflation in developing countries in Europe and Latin America and draws lessons from historical experience. It analyzes the dynamics of the interaction among deficit finance, institutional innovation in financial markets, dollarization, and the shortening of wage contracts in high-inflation situations. When stabilization is undertaken, there is neither immediate, spontaneous resumption of longer adjustment periods for wages and prices nor instant increase of real money demand to noninflationary levels. Incomes policy—freezing exchange rates, wages, and prices—is advocated as an effective supplement to the inevitable budget cut to make up for institutional inertia and facilitate the start of the stabilization process.

In early 1991 Argentina, Brazil, Peru, Poland, and Yugoslavia were in the midst of extreme instability or at best in the early stages of stabilization. Another group of countries, including the Soviet Union, Romania, and Bulgaria, were on the verge of slipping into high or even extreme inflation. A third group had already run the course and stabilized, as did Bolivia and Israel, or had avoided the extreme experience and opted for stabilization early and decidedly, as Mexico did.

The evidence from some 20 experiences with high inflation establishes that the similarities between the experiences of various countries become sharper and clearer and the differences less significant as the inflation rate rises. The particular mechanism by which monetary expansion occurs may differ—say, deficits of state enterprises rather than of a particular ministry—but the general pattern that runs from deficits to an expansion of money and credit is broadly the same, as are the dynamics of inflation. Of course the experience in the post-communist economies is special in that it starts from repressed inflation, but even that is not very different from experiences in Argentina or Brazil, where cycles of price controls and hyperinflation are now common (Cardoso 1991; Dornbusch, Sturzenegger, and Wolf 1990). The experience of Poland and Yugoslavia, and

Rudiger Dornbusch is Ford International Professor of Economics at the Massachusetts Institute of Technology.

the extraordinary problems already apparent in the Soviet Union, Bulgaria, and Romania, suggest that discussion of high inflation is timely (Commander and Coricelli 1992; Coricelli and Rocha 1990).

I. Lessons from History

Should hyperinflation make policymakers opt for zero inflation at any price? Or is there room in between, with cost-benefit analysis and with the lessons from a rich inflation experience across time and space? Three questions are of special interest in the context of inflationary instability and stabilization. First, how does a country fall into hyperinflation? Second, what is necessary to stop high inflation and return to normal growth? And, third, will Eastern Europe and the Soviet Union soon resemble Latin America, with some success stories, a few countries on the verge of high or extreme inflation, and some countries experiencing hyperinflation?

The statistics are, of course, open to question, but the size of the Soviet Union's budget deficits in the past few years indicate the problem ahead (table 1). The data show persistent and increasing deficits. Inflation is still negligible in official markets, but the stage is set for a dramatic inflation unless both the overhang and the deficit are addressed at the outset of any attempt to restructure the economy.

The answers to these questions can to a large extent be discerned from historical experiences of high inflation (Bruno and others 1988; Yeager 1981; Dornbusch and Fischer 1986; Dornbusch, Sturzenegger, and Wolf 1990; Dornbusch and Wolf 1990). Some of the lessons from the past are described below.

The Similarity of Inflationary Experiences across Countries

It is a mistake to believe that the problems of a particular country are unique. But it is common that policymakers in Brazil, Peru, or the Soviet Union cannot accept that the experience of their country is not unique in the essential facts of inflation. Yet their situation is not substantially different from that of the 20

Table 1. *The Soviet Budget Deficit, 1985–90*

	1985	*1986*	*1987*	*1988*	*1989*	*1990*[a]
Percentage of GNP						
Revenues	47.3	45.8	43.6	41.7	41.0	42.8
Spending	49.7	52.0	52.0	51.0	49.5	50.6
Budget deficit	2.4	6.2	8.4	9.2	8.5	7.9
Percentage per year						
Retail inflation	—	2.0	1.3	0.6	2.0	4.8
Money incomes	—	3.6	3.9	9.2	13.1	14.5

— Not available.

a. Plan and estimates.

Source: IMF, World Bank, OECD, and EBRD (1990).

other countries where policymakers also thought that their unique situation set ordinary economics aside (Dornbusch and Edwards 1991).

The Danger of Complacency

Hyperinflation is not around the corner whenever there is a budget deficit. But inflation can easily become a habit—and from there an unstable process. Complacency comes at a disastrous price: society falls apart as the middle class disappears and society is divided between those who know how to get ahead with inflation and those that fall behind (see Guttman and Meehan 1976 and Fergusson 1975 for dramatic descriptions of the German experience). Pauperization of the middle class rapidly corrodes social institutions. Public administration, the tax system, and all social relations become undermined by corruption and fraud. The middle class revolts against the state, and the poor revolt against property.

Participatory Democracy and Instability

Political change toward a more participatory democracy has not been the traditional vehicle for stability. Political change may carry with it the expectation of an improvement in opportunities and living standards. An inflationary response to raised expectations occurred in Europe in the 1920s, in the Soviet Union in 1919–21, during the Allende period in Chile, and with the growth of Solidarity in Poland.

Destructive inflation may accomplish many things that had been considered politically impossible. The destruction that takes place calls for particularly stern measures to rebuild confidence and stability. Because democratic institutions do not facilitate hard choices, democratic countries have almost invariably implemented special procedures to adopt and implement the hard measures necessary for stabilization. The arrangements differ from national unity governments (Israel in 1985) to restricted special powers for the executive (Poincare in France in 1926) or special parliamentary committees charged to interact expeditiously with the executive (Germany in 1923). In the end the politically impossible gets done because the destruction brought about by uncontrolled inflation is so devastating that it forces cooperation.

Fiscal Austerity

Stabilization does not imply that zero inflation must be achieved at any cost. Some policies incorporate a moderate rate of inflation because there is a steeply rising cost of disinflation. Policies that do not set a limit on inflation or policies that merely repress it for a while do not encourage confidence and stability. Policymakers need to have a good grasp of the role of control of the budget and incomes policy in stabilization. Without fiscal austerity stabilization cannot last; without incomes policy it cannot start.

Table 2. *Recent European Experience with High Inflation, 1986–90*
(percentage per year)

Year	*Hungary*	*Turkey*	*Poland*	*Yugoslavia*	*Soviet Union*
1986	5.2	34.6	17.7	89.8	2.0
1987	8.7	38.8	25.2	120.8	1.3
1988	15.6	75.4	60.0	194.1	0.6
1989	16.9	69.6	251.1	1,239.9	2.0
1990	28.3	60.3	585.8	583.1	4.8

Source: IMF, *International Financial Statistics* (various issues), and IMF, World Bank, OECD, and EBRD (1990).

Structural Reform and External Support

A separate set of issues concerns the transition from stabilization to growth. What kind of policies at home and abroad can help to decrease the risk of protracted stagnation? Structural reform and external support play a role in reassuring potential investors and thus moving the economy to growth (Dornbusch 1991).

II. Varieties of Inflationary Experience

Table 2 shows inflation rates for several European countries in the range from moderate to acute and extreme inflation. Table 3 shows average annual inflation rates in industrial and developing countries. High inflation is a problem of the developing countries of Europe and Latin America; it is not endemic to Africa, Asia, or industrial countries.

Extreme inflation (hyperinflation) is rare. The generally accepted operational definition of hyperinflation proposed by Cagan (1956) sets the benchmark at an inflation rate of 50 percent a month (12,875 percent at an annual rate):

> The term hyperinflation must be properly defined. I define hyperinflation as beginning in the month the rise in process exceeds 50 percent and as ending in the month before the monthly rise in prices drops below that amount and stays below for at least a year. The definition does not rule out

Table 3. *Inflation around the World, 1970–89*
(percentage per year)

		Developing countries				
Year	*Industrial countries*	*Africa*	*Asia*	*Europe*	*Middle East*	*Latin America and the Caribbean*
1970	5.6	5.4	6.5	—	3.0	12.4
1979	9.2	15.4	7.5	20.9	10.2	50.2
1988–89	4.7	23.3	11.8	126.0	17.0	208.0

— Not available.
Source: IMF, *International Financial Statistics* (various issues).

a rise in prices at a rate below 50 percent per month for the intervening months, and many of these months have rates below that figure.

Until recently there were few cases of hyperinflation in modern history. But now, with fresh cases emerging in Latin America and possibly in Eastern Europe, the phenomenon is more pervasive.

The distinction between hyperinflation and cases of lower and yet extreme inflation is somewhat arbitrary. Whether the inflation rate is really 50 percent a month or only 20 does not make too much difference because in either case inflation will be the dominant factor in the economy and will overshadow most other issues. Countries experiencing inflation rates of 10 or 15 percent a month for any length of time are moving toward hyperinflation. Table 4 shows the movement into high inflation in Argentina, Bolivia, Brazil, Mexico, Peru, and Israel in the 1980s. Unlike the inflation experiences associated with war dislocation or civil war, high inflation in these six countries is rooted in domestic mismanagement and, to some extent, in external shocks. In each case the transition to extreme high inflation started off with quite moderate rates, then suddenly gathered speed and became extreme. Only Mexico cut the process short and stabilized before all the mechanisms of instability could gather force.

Stabilization cannot afford to be weak. "Soft measures do not create hard currencies" the German authorities said in 1948 when a drastic monetary reform had to be administered (Dornbusch and Wolf 1990). The lack of thorough reform in Argentina, Brazil, and Peru shows up in the continuation for more than five years of off-and-on-again extreme inflation, which of course is accompanied by a dramatic decline in economic activity and the standard of living.

III. The Sources and Dynamics of High Inflation

This section discusses the interaction of financing requirements and the financial structure by assuming full wage-price flexibility. It focuses on the role of contracts in the inflation process and the dynamics of the interaction among

Table 4. *Recent Experiences with High Inflation, 1981–89*
(percentage per year, December to December)

Year	*Argentina*	*Bolivia*	*Brazil*	*Mexico*	*Peru*	*Israel*
1981	105	29	106	28	75	117
1982	165	133	98	59	64	120
1983	344	269	142	102	111	146
1984	627	1,281	197	66	110	374
1985	672	11,748	227	58	163	305
1986	91	276	145	86	78	48
1987	132	15	230	132	86	20
1988	343	16	682	114	10,205	16
1989	3,079	15	1,287	20	3,390	20
1990	2,314	17	2,938	27	7,482	16

Source: IMF, *International Financial Statistics* (various issues).

deficit finance, institutional innovation in financial markets, dollarization, and shortening contracts. Explosive inflation arises from the disintegration or melting of several institutions. A framework of the main determinants of inflation highlights the roles of budget finance, tax and financial institutions, and contracts in creating high inflation. The following analysis not only identifies the determinants of inflation but explains the mechanics of the very sharp acceleration that has been witnessed on several occasions.

Deficit Finance

There is considerable controversy in high-inflation countries about the exact, or even the approximate, size of budget deficits. Reliable public data, covering an extended period of time in a comparable fashion, are simply unavailable. Various series differ in their coverage of the public sector, in the distinction between budget and cash bases, and in the inclusion of certain expenditure items, especially with respect to the quasi-fiscal deficit of the central bank.

Both the adjustment of velocity and the presence of alternative means of financing the deficit (foreign borrowing, use of reserves, and domestic debt finance) help explain the lack of a tight link between inflation and the deficit. Controversies arise about the reason for the budget deficits, their endogeneity as a result of inflation, and their amplification by financial adaptation. That the actual outburst of inflation is often triggered by a foreign exchange crisis does not alter the fact that high inflation is a fiscal phenomenon.

The most common view asserts that high inflation is the result of budget deficits. If the government spends more than it receives in tax collection, the remainder is financed by creating money. That means more money—too much money—chasing too few goods with the predictable outcome of inflation. This view needs considerable refinement to be entirely correct. Three directions of correction are essential. First, there is some room for noninflationary deficit finance. Second, deficits can be financed by debt. Third, there is a channel of causation that runs from inflation to deficits, as well as the other way around.

A model of these important qualifications is given by equation 1. The deficit can be financed with high-powered money, with domestic debt, or with foreign debt:

$$gY = \dot{M}/P + \dot{B}/P + \dot{B}^{*} e/P \tag{1}$$

where g is the deficit ratio, Y is real gross domestic product (GDP), M is the domestic base money, B and B^* are domestic and foreign debt, e is the exchange rate (domestic currency over dollars), and P is the price level. (A dot over a variable denotes the rate of change.) It is clear that deficits can be financed by borrowing from abroad or at home, thus entirely avoiding an increase in the money stock, at least for the time being. Focusing on the situation in which the entire deficit is financed by money will show how much inflation can be generated by such a system.

Inflationary finance. Financing deficits by money creation means that any money that is not demanded at the current level of prices must be forced on the public by inflation. In a growing economy some extra real money balances are demanded in order to finance the growing level of transactions. But, beyond that, the demand for nominal money expands only to the extent that inflation erodes the purchasing power of existing real balances. To restore their real balances (at least partially), the public has to add to nominal money holdings. Thus inflationary finance automatically creates a demand for the money issue that finances the deficit.

Keynes (1923, p. 37), in his splendid description of the inflation tax, noted the scope for inflationary finance even in a country with the poorest economic and political conditions:

> A government can live for a long time, even the German government or the Russian government, by printing paper money. That is to say, it can by this means secure the command over real resources, resources just as real as those obtained by taxation. The method is condemned, but its efficacy, up to a point, must be admitted. . . . so long as the public use money at all, the government can continue to raise resources by inflation . . . a government can get resources by a continuous practice of inflation, even when this is foreseen by the public generally, unless the sums they seek to raise in this way are very grossly excessive. . . . What is raised by printing notes is just as much taken from the public as is a beer duty or an income tax. What a government spends the public pays for. There is no such thing as an uncovered deficit.

But, as Keynes has noted, significant inflation reduces the amount of money people choose to hold, because they will substitute toward assets that are more inflation proof. Thus, just as high taxation erodes the tax base, high inflation leads to a reduction in real balances and hence to an increase in the rate of inflation necessary to finance a given deficit. Moreover there may be a maximum amount of resources the government can extract.

The long-run relation between the money-financed budget deficit and the rate of inflation is shown in equation 2 (for a derivation, see Dornbusch 1985):

$$\pi = (\alpha g - y)/(1 - \beta g),\ 1 > \beta g \qquad (2)$$

where π and y are the rate of inflation and the growth rate of real GDP. The term α represents the noninflationary level of velocity, and β is the responsiveness of velocity to the rate of inflation. This equation shows that because the deficit is financed by money creation, there is inflation. But it also shows that the inflationary impact of a given deficit can differ widely, depending on the financial structure and the growth rate of output.

There are three key points of this relation. First, the inflation rate is lower the higher the growth rate of output is. When output grows strongly, so does the

demand for real money. Accordingly there is room for some extra money issued without introducing the risk of inflation. Second, inflation is higher the larger the budget deficit is. Moreover this relation is nonlinear. As the government tries to finance a larger deficit, the required rate of inflation increases steeply. Depending on the particular form of the money demand equation, there may even be a maximum deficit that can be financed by money. Going beyond that range implies hyperinflation. Third, the inflation rate depends on the velocity parameters in equation 2. The higher is the level of noninflationary velocity (that is, because of dollarization) the higher the rate of inflation associated with any given deficit. A high degree of responsiveness of velocity to inflation also implies a larger rate of inflation.

The increase in inflation brought about by a one percentage point increase in the deficit is higher, the higher are inflation and the budget deficit from which one starts. Inflationary finance thus exerts a very powerful impact on inflation if it is used in large doses or in an environment where a high level of velocity, and strong responsiveness of velocity to inflation, leaves little scope for an inflation tax. Likewise, dollarization or a drop in growth bring about large increases in the inflation rate, more so the higher the initial extent of deficit finance.

The Olivera-Tanzi effect. One of the striking effects of inflation is the erosion of the real value of taxation. If there is any delay between accrual and payment of taxes, the inflation in the interim will mean that the real value of what is paid is lower the higher the rate of inflation. With moderate inflation it makes no difference that 1987 taxes are paid in 1988. But when inflation is high, this effect wreaks havoc with the real value of tax collection. Keynes, commenting on the impact of inflation on the budget noted this point, as did Bresciani-Turroni (1937). Tanzi (1978) and others have recognized this effect in the specific context of Latin American inflation. The empirical importance of this effect is large whenever inflation is high and tax collection lags are long and when there is no provision for tax indexation.

External shocks and inflation. Suppose, as is the case in Argentina, that the public sector has a large external debt and an external debt shock occurs. Specifically, assume that before the disturbance any existing external debt was rolled over with interest fully capitalized through automatic "new money" and that there is no domestic debt. Let d^* be the flow of external debt service (measured as a percent of gross national product [GNP]), and thus g is the total deficit ratio that is financed by money creation. Thus,

$$g = \sigma(\pi) + d^* \tag{3}$$

From equation 3 reduced access to automatic capitalization of interest payments implies that external debt service leads to increased deficit finance by money creation. The country has to earn the resources for external finance or else finance the purchase of foreign exchange by creating money. First, the

ment will issue more money to finance the purchase of foreign exchange interest payments (assuming, of course, that there are no expenditure cuts or increases). Second, there will typically be a real depreciation in order to prove the external balance.

The increase in inflation resulting from an external financing disruption is ger, the larger is the debt service shock and the real depreciation, but it also pends on the responsiveness of velocity to inflation and on the degree to which creased inflation erodes real tax collection. Each of these factors will increase e inflationary impact of the debt shock significantly.

The "balance of payments school" would argue that external balance problems and the resulting depreciation of the exchange rate are the primary causes f the deficit. By contrast, the "quantity theory school" would point to budget deficits and their financing by money creation as the reason for inflation. Passive money is the essential ingredient in reconciling the quantity school and the balance of payments doctrine. Not surprisingly, suspension of reparation payments in Germany and of debt service in Bolivia in 1985 were essential steps in the stabilization of inflation. In Argentina involuntary external debt service after 1982 became an important source of inflation in exactly the manner the balance of payments school emphasizes. Deteriorating terms of trade further aggravated the external debt shock by forcing real depreciation and hence an increase in the real value (in terms of GDP or the tax base) of the existing external debt service.

Endogenous financial innovation and liberalization. The interest that traditional depository institutions can pay is typically controlled. There may be an outright limitation on interest rates, or else institutions may be required to hold reserves or government debt at controlled rates. These restrictions make institutions unable to compete in financial markets where nominal interest rates more nearly reflect the ongoing inflation. New, unregulated financial institutions that offer depositors higher interest rates spring up and thus draw customers away from traditional depository institutions. There is a fall in the ratio of conventional money to GDP. The government loses part of its inflation tax base, and hence equilibrium inflation increases. The government may aggravate matters when it responds to the increasing inflation by raising reserve requirements or forcing traditional banks to hold government debt.

Governments often promote this process, most obviously under the guise of financial liberalization. Since inflation is a tax on money (or commercial bank non-interest-bearing reserves), financial liberalization means that the public can avoid the tax on money. Financial liberalization may take the form of interest-bearing deposits or formal dollarization, each of which reduces the demand for high-powered money; velocity rises and so does the inflation rate associated with the financing of a given deficit by money creation. Thus, from an inflation point of view, financial repression, not liberalization, is appropriate. Financial liberalization requires that extra tax revenue be available to avoid the inflationary impact of a reduction in the captive inflation tax base. Governments that

condone dollarization likewise promote inflation. Dollarization is captured in equation 2 by both the coefficients α and β. The shift from the domestic monetary base into dollars reduces the base for the inflation tax and hence must increase inflation.

One is tempted to explain inflation experiences in some countries by dollarization and new financial intermediaries. Thus countries with stronger dollarization have higher inflation. A government that experiences some inflation and makes dollarization easier will experience even more inflation. However, dollarization is also a response to inflation.

The financial adaptation to inflation intensifies the inflationary process. In response to inflation there is a flight from money into interest-bearing financial assets, to the extent that they exist at all, or into dollars. But there is also an institutional adaptation: financial institutions spring up that offer protection against inflation. The better the protection they offer, the more substantial the flight from money or the larger the increase in velocity.

Timing. Inflation controls and managed exchange rates can slow down the buildup of inflation. The loss of resources, or forced saving, is an alternative mode of financing. However, deficits do imply money creation and inflation. Moreover the longer the delay the more dramatic the inflationary explosion. This is especially the case when a managed exchange rate and reserve losses have financed the deficit in a relatively noninflationary manner. When these mechanisms are no longer possible, there will be a sudden shift toward the inflation tax at the same time that a real depreciation is required.

The Role of Contracts in the Inflation Process

As inflation accelerates, contracts shorten, and that shortening of contracts is itself a factor that causes inflation to accelerate. Institutional wage-setting mechanisms often rely on a fixed contract length, with wage adjustments occurring at specified intervals. The adjustments are based on the cumulated increase in prices since the last adjustment. For example, earners might receive full compensation for past actual price increases at regular intervals, say yearly. Now suppose there is a shift to six-month intervals. There are two interesting questions. The first concerns the dynamics of shifting to shorter contracts. What is the threshold for inflationary erosion of wages that causes the shift, and what makes it economywide rather than just for a particular firm? The other interesting question is what happens when the frequency of adjustment increases yet further. This point has been developed especially by Pazos (1972). It is of interest here because contract deterioration is one of the important characteristics of an accelerating inflation and because exchange depreciation often plays an important role in setting off the process.

If nominal wages are adjusted only periodically, the real wage follows a sawtooth pattern. On each adjustment date the nominal wage is increased by the cumulated inflation since the preceding adjustment. Until the next adjustment

ate the real wage declines as the ongoing inflation erodes the purchasing power f the constant nominal payments. By the end of the adjustment interval the real rage has declined below its period average. The higher the rate of inflation, ıoreover, the lower the average real wage, given the interval of adjustment.

In a system of full, but lagged, indexation, the real wage can be cut only by ıoving to a higher rate of inflation. Thus, once-and-for-all depreciation of the urrency immediately raises the rate of inflation and erodes existing contracts. ut wage indexation ensures that inflation must be pushed to an even higher rate › that there is always some group of wage earners whose wages are still lagging ıe increasing rates of price increases. The same principle applies to the removal f subsidies undertaken to correct the budget. Measures undertaken to correct ɔmpetitiveness or the budget can be effective only if they achieve a cut in the :al wage, but because of full indexation that cut can take place only if inflation allowed to run at a higher rate. This mechanism often sets the stage for ıflation explosions.

Consider a country that requires adjustments in the budget and external competitiveness. Suppose that the government lacks the political force to suspend full indexation, so that the removal of subsidies or a real exchange depreciation will speed up the inflation rate. Workers in the middle of their contracts, for example, will find that their real wages fall below what they consider a minimum standard of living. They cannot borrow, even in perfect capital markets. Hence they will call for a shorter interval between wage adjustments in order to recover the real wage losses imposed by inflation. They will ask for an advance of what they think is due. If the economy does, in fact, shift from, say, six-month to three-month indexation intervals, the inflation rate will simply double (Simonsen 1986). But once the contract structure has moved to a three-month scheme, it is unlikely that the indexation structure will return spontaneously to a onger interval, even if shocks are favorable. And there is nothing to make the three-month interval more stable than the six-month interval that was just abandoned. New shocks will shift the economy to even more frequent adjustments and hence to correspondingly higher rates of inflation. At this stage the exchange rate becomes critical.

The dramatic escalation of inflation, seemingly out of proportion to the disturbances, arises from the endogeneity of the adjustment interval. This is due not so much to the direct impact on inflation of corrective exchange rate or price policies. It occurs because increases in inflation, which may be minor but highly visible (such as a 10 percent devaluation over and above a purchasing power parity rule or a removal of bread subsidies), lead to an increase in the frequency of wage adjustments, which brings on a much higher inflation rate. The endogeneity of adjustment intervals is the mechanism that connects small inflation disturbances with a shift from 50 to 100 percent inflation or beyond to hyperinflation. As long as full indexation remains, even seemingly small corrections are a dramatic threat to the stability of the inflation rate and hence may not be worth undertaking. Incomes policy designed to avoid inflationary explosion must avoid accelerating the frequency of adjustments.

Dynamics

The actual dynamics of the economy emerge from the interaction of the inflationary aspects of deficit finance and the contracting process. A stable equilibrium may not actually exist. When inflation rises significantly and permanently, institutions adapt. In doing so, they help to increase inflation. Under conditions of extreme inflation, institutions break down. There is a near-abandonment of domestic money, which means the government must continue to increase inflation to get any seigniorage. Contracts are set for a shorter duration and are more likely to be dollar-based.

In the analysis of inflationary experiences, it is common to assume adaptive inflationary expectations (Cagan 1956). There appears to be a significant sluggishness in the initial phases of high inflation as well as a subsequent acceleration, which suggests exactly such an expectations mechanism. Adaptive inflationary expectations are often the key model device to slowing the impact of money on inflation. An alternative and perhaps more accurate model focuses much more on the dynamics of deterioration in contracts, both in the goods and labor markets, and on the inflationary adaptation of financial institutions. Institutional dynamics seem to offer a more suitable framework for studying high inflation.

As economic institutions break down and time intervals for contracts and adaptation to inflation become shorter, the inflation process becomes explosive. The economic time horizon shrinks along with contracts and maturities of financial assets until, when the economy converges to a spot market with dollar pricing, the budget or external balance deficit leads to hyperinflation. Hyperinflation is inevitable because the inflation tax, with sufficient financial adaptation, can be almost totally evaded, and hence the budget deficit cannot be financed. The Olivera-Tanzi effect, the shortening of contracts, and financial adaptation all react in a perverse way (from the perspective of stabilization) in that they widen the deficit and accelerate explosively the inflation process.

IV. Stabilization

The preceding discussion helps to explain why stabilization is difficult and often takes more than one attempt to succeed. In the process of high inflation all institutions break down. When stabilization is undertaken, there is neither immediate, spontaneous resumption of longer adjustment periods for wages and prices nor an instant increase of real money demand to noninflationary levels. As a result more sizable adjustments in the budget are required, and more dramatic measures are necessary to create the confidence that stabilization will, in fact, last. Because the fiscal measures have to be particularly large they are also particularly difficult and hence often can not be sustained. When they fail, inflation returns instantly at exceptionally high levels because institutional inertia had not recovered.

Incomes policy—freezing exchange rates, wages, and prices—can be an effec-
ive supplement to the inevitable budget cut. It makes up for institutional inertia
nd, to that extent, gives a government a better chance to start stabilization.
ut, as is clear from the experiences of Argentina, Brazil, and Peru, failure to
orrect the budget implies that high inflation will soon return. The decline in the
atio of M1 to GDP is not typically fully reversed in the initial stabilization. As a
esult financing even a moderate deficit is much more inflationary than it was
efore the experience of extremely high inflation. This hysteresis effect of high
nflation (similarly apparent in contracts, pricing, and tax collection) sharply
educes the chances of stopping inflation with anything short of a dramatic
udget cut.

The task of stabilizing inflation involves stopping inflation quickly and avoid-
ng the resurgence of inflationary pressures. To end inflation by incomes policy is
elatively easy, but to keep it down requires fiscal support. The chief mistake in
abilization policy is to rely too much on incomes policy—fixed exchange rates
nd wage and price freezes—and too little on fiscal austerity. Such programs
uickly lead to repressed inflation and overvaluation, in which tight monetary
olicy is introduced to sustain the imbalances. Ultimately that does not work,
d another inflationary explosion offers the starting point for yet another
abilization. Argentina offers a clear example of this process with its successive
iled stabilization programs in the past five years.

Budget Balancing

Budget deficits are the ultimate source of inflation. When external financing
the domestic capital market cannot finance deficits, then the deficits must be
justed. Two questions immediately emerge. The first is how large a deficit is
nsistent with stability; the second is how to cut deficits down to the required
e.

Argentina, Brazil, and Peru failed to adjust fiscal deficits in the aftermath of
eir 1985 heterodox stabilization. Wage-price controls and fixed exchange
tes quickly stopped inflation and raised the political popularity of the presi-
nt. The resulting possibility for fiscal stabilization was, however, not used.
stead the deficits persisted and were financed by creating money. As a result,
flation continues.

Quasi-fiscal deficits. The starting point for budget balancing is the need for a
nsparent accounting of the consolidated government. Because the issue is
ntrol of monetary emission, it is essential that the central bank's "quasi-fiscal"
ficit be part of the accounting. An accounting framework is needed for the
nsolidated government sector, including not only the central government and
e central bank, but also state enterprises and local government. Extreme
flation invariably reflects deficits financed by writing checks on the central
nk, whether it be by provincial authorities, as in China; by state enterprises,
in Yugoslavia; or by a government bank, as in Brazil. The deficits may have as

a counterpart purchases of foreign exchange, payments of wages, deficits of railroads, external debt service, election spending by a governor, or sim corruption. In any case a deficit leads to money creation.

Quasi-fiscal deficits arise from loans by the central bank at subsidized rat losses on foreign exchange operations in the form of guarantees, forward c tracts, or simply purchases at a high rate (under multiple exchange rates) a sales at a low one. In Peru in 1986–87, for example, exchange losses accoun for 2.3 percent of GDP. But central bank losses also arise from credit operatio Subsidized credit is no different from any other subsidy; in fact credit subsic have long ceased being investment subsidies and have become simply a prod tion subsidy that finances wages when prices are not allowed to reflect costs.

Revenues. The second point on the reform agenda is to achieve a product tax system. The reform must raise revenue on a substantially larger scale and more efficiently. Increasing the yield of the tax system is dictated by the need eliminate deficits. Inflation stabilization makes an immediate contribution cause the inflationary erosion of revenues ceases. But that is only a small pa perhaps as much as 2 percent of GNP in revenues. The major effort must be reconstructing the tax system, including stopping the corruption and evas that now undermine the collection of taxes as well as introducing and dem strating mechanisms to increase compliance. The complacent acceptance of p vasive tax evasion is the most regressive aspect of the Latin American tax s tem, and it must be carefully watched as Eastern Europe moves to taxation the chief source of government revenue.

The revenue effort must concentrate both on collecting taxes and on elimin ing subsidies in public sector enterprises. Many countries now have pervas systems to manage public sector prices, both to control inflation and to try prevent a decline in real wages. The implied revenue losses are extraordina large and cannot be justified by any of the objectives. For example, in P controlled telephone rates have reduced the real price of the service to one-te the 1985 level. It is difficult to argue that telephone rates have an import incidence either on inflation or on welfare of the poor, but they do contribute deficits.

Governments should therefore eliminate totally all subsidies. The result revenue gains must be applied to eliminate inflation, which in itself raises welfare of the poor since inflation is a highly regressive tax. Part of the reven should also be used for targeted food and employment programs for the poor groups.

Beyond cutting all subsidies and raising revenue under the existing structu governments should use the crisis to institute a more efficient tax system. T system should produce more revenue with fewer distortions, which means eli inating the pervasive exemptions from direct taxes and raising the rates to hig levels. A comprehensive value added tax of 15 percent, with a 5 percent s charge for luxuries, might be the starting point for discussion.

Government spending. For many observers the right direction for adjustment is to cut government spending, not to raise taxes. Inefficiency in government is pervasive, and public sector employment in many countries is unjustifiably high. But there is no presumption that the fiscal problem can be solved correctly by massive firing of public sector employees and privatization. There is a need to restructure public sector spending, from consumption to investment and productive services. But as for the level of spending, it certainly is not excessive. More of the spending absolutely and relatively should fall on infrastructure, health, and social services for the poorer groups. The current composition of spending is not only unproductive but probably also regressive.

Most of infrastructure investment could be done by the private sector. That is certainly the case, for example, for telephone services, but also for public transport and even the road system. Mexico is now exploring such options with very substantial success. But the fact remains that infrastructure spending should not be the priority in balancing the budget, certainly not at the expense of health and education.

Incomes Policy

Fiscal austerity is the essential aspect of stabilization, but incomes policy is an important, desirable component. Incomes policy is designed to bring about a rapid, coordinated end to inflation. In a hyperinflation, incomes policy amounts to fixing the exchange rate. Because price setting is geared to the movements of the dollar, the move to a fixed exchange rate is enough to break the inflation and the expectation of inflation.

But when annual inflation is only 100 or 200 percent, incomes policy is both more essential and more complicated. Without incomes policy ending inflation by demand management alone would create an extraordinary depression. The current inflation will be a weighted average of cost increases that are equal to past inflation, which enters costs by explicit or implicit indexation, and the current rate of exchange depreciation, $\dot{e}$, plus a cyclical component, which is denoted by "GAP":

$$\pi = \tau\pi_{-1} + (1-\tau)\,\dot{e} + \Psi \text{GAP}. \tag{4}$$

Because of the inertia represented by the cost increases resulting from explicit or implicit indexation, current inflation cannot get away from past inflation unless the government breaks the process by incomes policy. Incomes policy means fixing the exchange rate and stopping wage inflation. The government will have to intervene in loan contracts to reduce real interest burdens that otherwise would result from the unanticipated decline in inflation, and intervention will be required in wage contracts. Because these contracts have periodic adjustments for inflationary erosion, a sudden ending of inflation requires intervention. Some wage contracts have to be rolled back, and others need to be adjusted upward.

Exchange rate policy. Exchange rate policy assumes a strategic role in stabilization, as does pricing in the public sector. The starting point of a program is invariably a fixed exchange rate. But if inflation does not end completely, sooner or later adjustments in the exchange rate and public sector prices are needed. The decision to abandon the fixed rate is a difficult one because it signals the government's acceptance of inflation as something inevitable. As a result there is a temptation to postpone exchange rate adjustment until a significant overvaluation has developed.

Overvaluation in turn creates an expectation of a devaluation, and very high real interest rates become necessary to stop speculation. High real interest rates in turn increase domestic debt service and worsen the budget. Ultimately the exchange rate adjustment does have to come, but often the overvaluation has gone so far that an outright exchange crisis and collapse are the end of the abortive attempt to practice a fixed exchange rate. The pragmatic answer is to move after two or three months to a crawling peg, depreciating the exchange rate at a pace that maintains external competitiveness. The risk of an overvaluation maintains short economic horizons and stands in the way of recovery. The right time for a crawling peg is very early because the government should try to preserve maximum competitiveness. Holding onto an exchange rate too long may yield an extra month of low inflation, but it also sacrifices competitiveness and therefore prejudices the return of growth.

Indexation. A major stabilization decision regards indexation. The common view is that indexation is responsible for the inflation and that accordingly it should be abolished. Moreover governments should declare a zero inflation target rather than create mechanisms that make it easier to live with inflation. However, it is not necessarily true that without indexation there is inflation stability because inflationary shocks, such as public sector price increases and depreciation, are not fully absorbed into lower real wages. Without explicit indexation the government becomes the judge of what wage increases to grant. The wage becomes politicized, which means invariably larger rather than smaller wage increases, and wage increases come sooner rather than later. In fact in economies in which a government seeks to avoid explicit indexation, as in Brazil after 1985, inflation soon becomes more unstable and susceptible to a far more rapid escalation than had ever been experienced under indexation.

Indexation is a mechanism that creates inertia and also preserves inertia. Reintroducing half-yearly indexation may therefore be a key step in establishing the expectation of low inflation. Once the wage is locked away, a very rapid resumption of inflation will not be expected. As a result horizons can lengthen far more effectively than under threshold provisions or in the absence of any kind of formal indexation.

Monetary policy. Monetary policy does not play an independent role in stabilization; it is dictated by the budget and the exchange rate policy. Following

stabilization real interest rates are too high. One could argue that to resume growth the economy needs reliquification. There is very limited room for reliquification, but that is best done by monetizing reserve inflows rather than by deficit finance or domestic credit creation.

The alternative is an overly firm commitment to a zero inflation target. The policymaker might be tempted to make monetary policy (and the exchange rate) do what fiscal policy has not achieved. The risk is a long period of extraordinarily high real interest rates and possibly an exchange overvaluation. They might stop inflation, but they also will destroy the real economy. The best and only lasting way to bring about low real interest rates and to achieve moderate inflation is by a balanced budget (including state enterprises, except where they are financed in the capital market) and a very competitive real exchange rate.

V. Concluding Remarks: Priorities

Without financial stability, economic reconstruction and growth will simply not occur. If inflation is high and variable, then it will be the most important issue. It will take up policymakers' precious time. It will tempt them into superficial remedies, which help in the short run but set back economic activity because they create uncertainty. It will also lead the private sector to focus on protection against inflation and government's arbitrary interventions. The first priority then must be to reduce inflation. A lastingly balanced budget is required to achieve financial stability.

In economies in which price control has been the rule, the recognition of the problem of inflation may be the most serious issue. There has been no experience with inflation, and most attention focuses on the popular revolt against removing some subsidies. But beyond the initial correction of subsidies there is the broader issue of the risk of a serious inflation. Serious inflation can emerge either because there is an initial monetary overhang or because the subsidy correction does not go far enough. Deficits remain, and money creation starts to interact with corrective inflation. (The Soviet Union is an obvious case in point.)

In countries with major fiscal problems it is not politically impossible to make the necessary adjustments; it is politically difficult, but the adjustments will ultimately be made. The only question is how large the loss in the standard of living has to be before it is done and how much time and political capital will be lost (Blanchard and others, 1991; Fischer and Gelb 1990, and Dornbusch 1990).

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Bresciani-Turroni, C. 1937. *The Economics of Inflation*. London: Allen & Unwin.

Blanchard, Oliver, Rudiger Dornbusch, Richard Layard, and Lawrence Summers. 1991. *Economic Reform in the East*. Cambridge, Mass.: MIT Press.

Bruno, Michael, Guido Di Tella, Rudiger Dornbusch, and Stanley Fischer, eds. 1988. *Stopping High Inflation*. Cambridge, Mass.: MIT Press.

Cagan, Phillip. 1956. "The Monetary Dynamics of Hyperinflation." In Milton Friedman, ed., *Studies in the Quantity Theory of Money*. University of Chicago Press.

Cardoso, Eliana. 1991. "From Inertia to Megainflation: Brazil's Macroeconomic Policies in the 1980s." In Michael Bruno and Stanley Fischer, eds., *Lessons of Economic Stabilization and Its Aftermath*. Cambridge, Mass.: MIT Press.

Commander, Simon, and Fabrizio Coricelli. 1992. "Price-Wage Dynamics and Inflation in Socialist Economies: Empirical Models for Hungary and Poland." *The World Bank Economic Review*, this issue.

Coricelli, Fabrizio, and Roberto Rocha. 1990. "Stabilization Programs in Eastern Europe: A Comparative Analysis of the Polish and Yugoslav Programs of 1990." World Bank, Washington, D.C. Processed.

Dornbusch, Rudiger. 1985. "Stopping Hyperinflation: Lessons from the German Experience in the 1920s." In Rudiger Dornbusch, Stanley Fischer, and John Bossons, eds., *Macroeconomics and Finance: Essays in Honor of Franco Modigliani*. Cambridge, Mass.: MIT Press.

———. 1990. "Priorities of Economic Reform in Eastern Europe and the Soviet Union." Massachusetts Institute of Technology, Department of Economics, Cambridge, Mass. Processed.

———. 1991. "Policies to Move from Stabilization to Growth." *Proceedings of the World Bank Annual Conference on Development Economics 1990*. Washington, D.C.: World Bank.

Dornbusch, Rudiger, and Sebastian Edwards. 1990. "The Macroeconomics of Populism in Latin America." *Trimestre Economico* 57 (January–March): 121–62.

———, eds. 1991. *The Macroeconomics of Populism in Latin America*. University of Chicago Press.

Dornbusch, Rudiger, and Stanley Fischer. 1986. "Stopping Hyperinflations: Past and Present." *Weltwirtschaftliches Archiv* 122 (1, April): 1–14.

Dornbusch, Rudiger, Federico Sturzenegger, and Holger Wolf. 1990. "Extreme Inflation: Dynamics and Stabilization." *Brookings Papers on Economic Activity* 2: 1–64.

Dornbusch, Rudiger, and Holger Wolf. 1990. "Monetary Overhang and Reforms in the 1940s." Massachusetts Institute of Technology, Department of Economics, Cambridge, Mass. Processed.

Fergusson, A. 1975. *When Money Dies*. London: William Kimber.

Fischer, Stanley, and Alan Gelb. 1990. "Issues in Socialist Economy Reform." Massachusetts Institute of Technology, Department of Economics, Cambridge, Mass. Processed.

Guttman, William, and P. Meehan. 1976. *The Great Inflation*. London: Gordon and Cremonesi.

International Monetary Fund. Various issues. *International Financial Statistics*. Washington, D.C.

International Monetary Fund, World Bank, Organisation for Economic Co-operation and Development, and European Bank for Reconstruction and Development. 1990. *The Economy of the Soviet Union*. Washington D.C.

Keynes, John M. 1923. *A Tract on Monetary Reform*. Reprinted by the Royal Economic Society, London, 1971.

Pazos, Felipe. 1972. *Chronic Inflation in Latin America*. New York: Praeger.

Simonsen, Mario. 1986. "Indexation: Current Theory and the Brazilian Experience." In Rudiger Dornbusch and Mario Simonsen, eds., *Inflation, Debt and Indexation*. Cambridge, Mass.: MIT Press.

Tanzi, Vito. 1978. "Inflation, Real Tax Revenue, and the Case for Inflationary Finance: Theory with an Application to Argentina." *IMF Staff Papers* 25 (September): 417–51.

Yeager, Leland. 1981. *Experiences with Stopping Inflation*. Washington D.C.: American Enterprise Institute.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 33-53

Price-Wage Dynamics and Inflation in Socialist Economies: Empirical Models for Hungary and Poland

Simon Commander and Fabrizio Coricelli

This article analyzes the determinants of open inflation in transitional socialist economies, with reference to recent experience in Hungary and Poland. A simple inflation model is centered on the transmission process and on the short-run dynamics of inflation. Further incorporating a number of features specific to socialist economies and working with quarterly data, dynamic price and wage equations are estimated. The estimated equations allow satisfactory exploration of the role and weight of foreign prices and domestic factors in propagating inflation. Foreign prices matter, but developments on the cost side are critical in relating exogenous, policy-driven adjustments to the price level to increases in the rate of inflation. The absence of conventional market-based, equilibrating mechanisms requires that nominal anchors, particularly wage restraints, feature prominently in any stabilization program adopted by reforming socialist economies.

An apparently unique feature of classical centrally planned economies has been the absence of inflation. However, it is generally accepted that inflation was repressed in centrally planned economies, with the result that open inflation has become a hallmark of economies making the transition to a market-based system. By itself the translation of repressed inflation into open inflation cannot explain the acceleration—and the persistently higher rates—of inflation that has been observed in reforming socialist economies. Moreover the rapidity of the acceleration threatens to dilute the desired signaling role for prices and jeopardizes the political sustainability of reform.

Although an extensive literature exists that deals with the measurement of repressed inflation (Nuti 1986; Portes 1977; Feltenstein 1989), analysis of open inflation is still largely anecdotal (Honohan 1989 for China and Rocha 1990 for Yugoslavia are among the few exceptions). Little attempt has been made to pin down empirically the consequences of the particular behavioral rules and institutional routines, particularly cost-plus pricing, that are generally agreed to be characteristic of socialist economies in transition. In this article we try to isolate

Simon Commander is with the Economic Development Institute, and Fabrizio Coricelli is with the Country Economics Department, both at the World Bank.

key empirical regularities of the inflation process in reforming socialist countries.

Several features of the socialist economies distinguish them from market-based economies and indicate areas for specific analytical treatment. Such disparities include the administrative determination of prices, the wage-setting framework and the imposition of full employment, external trading prices and regimes, and the monopolistic structure of goods markets. Further, such economies might be expected to lack any endogenous mechanism of equilibration, given full employment and an absence of a conventional relation between output and prices. Several of these features are incorporated in an inflation model, calling on Hungarian and Polish experience and data. To highlight the short-run dynamics and propagation mechanism, price and wage models are estimated using quarterly data that cover much of the recent reform period. The analysis pays particular attention to the relation between prices and wages and to the role of the exchange rate in determining domestic inflation.

Section I discusses the main institutional features and developments for prices and wages in Hungary and Poland. A simple structural price and wage model suitable for estimation is summarized in section II; the full model is laid out in the appendix. The reduced-form price and wage equations are estimated based on available quarterly data, with the results reported in section III. Section IV presents concluding remarks.

I. Institutional Features and Prices and Wages in Hungary and Poland

There are several institutional features related to prices and wages that are similar in Hungary and Poland. However, the inflationary experiences and the effect of wage determination on inflation have differed in the two countries.

Main Features of Price Reform in Hungary and Poland

Three main features can be extracted from Hungarian and Polish experience with price reform and inflation, especially during the 1980s. First, there has been a sustained upward shift in the price level. In the Hungarian case inflation has consistently ratcheted upward, exceeding 30 percent a year in 1990 with likely expansion to more than 50 percent in 1991. In the 1980s in Poland inflation remained consistently higher and exhibited far greater variance than in Hungary. This culminated in a high inflation burst in 1989, followed by a radical stabilization, which, by late 1990, brought inflation down to below 5 percent a month.

Second, both economies have moved away from strictly administratively determined prices. A series of price liberalization measures has reduced administered prices as a percentage of total prices. Table 1 provides information on the distribution of different price categories. However, neither free prices in Hungary nor contract prices in Poland have been wholly free from administrative interference. Established ratios of costs to profits in the material sector and

ble 1. *Consumer and Producer Prices by Pricing Category*
Hungary and Poland, 1982–90

untry and ice category	*1982*	*1984*	*1985*	*1987*	*1988*	*1989*	*1990*
ungary							
onsumer prices							
Fixed	15	15	6	6	0	0	n.a.
Flexible	29	28	38	32	20	17	n.a.
Free	56	57	56	62	80	83	n.a.
oducer prices							
Fixed	33[a]	30	32	28	25	23	n.a.
Free	67[a]	70	68	72	75	77	n.a.
oland							
onsumer prices[b]							
Administrative	35	47	47	45	45	45	10
Regulated	15	3	3	2	0	0	0
Contract	50	50	50	53	55	55	90

n.a. Not available.

Note: Data were not available for 1983 and 1986. Flexible consumer prices include those subject to naximum limit; fixed producer prices also include all prices with upper limits.

a. Data for 1980.

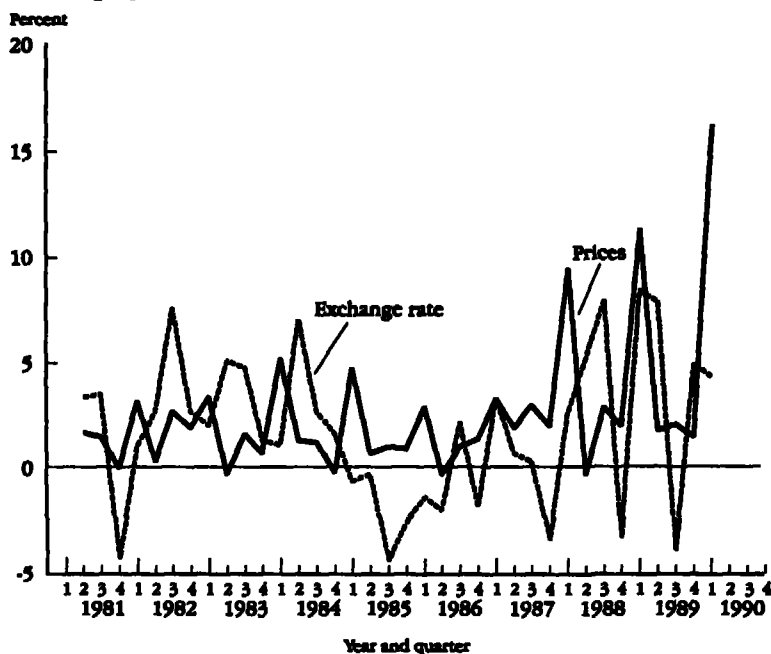
b. Expressed as share of total sales value for products covered by price category. Since 1990 changes in ontract prices no longer require notification.

Source: Data provided by the Ministries of Finance of Hungary and Poland.

ither guidelines have constrained price growth. In both economies this has iluted the potential impact on relative prices that might have been expected to esult from a shift in the locus and timing of price adjustments. Consequently, hajor relative price rearrangements did not occur in Hungary until after 1987; n Poland significant change took place in 1981–82, 1988–89, and, more radially, with the 1990 stabilization. In the 1990 stabilization program in Poland rice controls were almost totally abandoned.

Third, price liberalization has been associated with a greater explicit linking f domestic prices to foreign prices. In Hungary between 1973 and 1978 conumer prices expanded at least six percentage points below the rates of countries f the Organisation for Economic Co-operation and Development. In 1979–80 he price level was adjusted sharply upward to eliminate the wedge between roducer and consumer prices. Producer prices—principally for exportables—ere then explicitly linked to international prices through the competitive pricng rule. In Poland the impact of foreign prices on administered prices remained elatively weak. The equalization settlements system equalized transaction rices for exporters and importers to domestic market prices through subsidies nd taxes.

Figures 1 and 2 show that imported inflation has become progressively more owerful. In Hungary there was a pattern of staggered adjustment of producer nd consumer prices to foreign prices from the mid-1970s onward. Import and omestic consumer prices moved more closely together during the 1980s. After 986 Hungarian inflation began to exceed that of its convertible trading part-

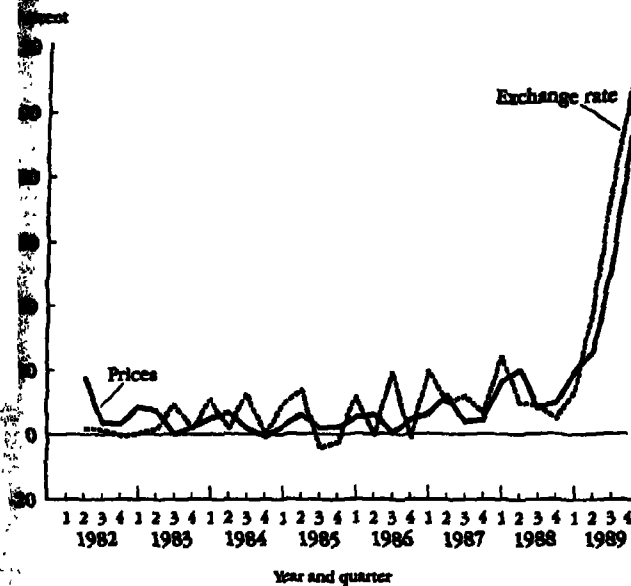
Figure 1. *Changes in Prices and the Official Exchange Rate in Hungary, 1981.2–90.1*

Source: International Monetary Fund (various years).

ners. In Poland import and domestic prices moved together after 1982. Implementing a competitive exchange rate policy aimed at boosting exports and reducing the gap between official and parallel rates resulted in recurrent depreciation and a strong imported inflation effect. Reduction in the share of administered prices in total prices and the explicit association between domestic and foreign prices appear to have allowed some linkage, albeit very weak, of price changes to the real side of the economy.

Inflationary Experiences under Reform

In both Hungary and Poland in the 1980s consumer prices, wages, and unit labor costs moved together closely (figures 3 and 4). However, similar ground rules for reform generated radically divergent inflationary paths. Thus the respective inflationary experiences in the two countries must be analyzed in relation to the conditions under which the accelerated reforms of the 1980s were enacted.

Inflation in Hungary. An annual inflation model covering 1960–87 and relating change in consumer prices to lagged inflation, import prices, unit labor costs, and an excess demand term provides a reasonably robust explanation of

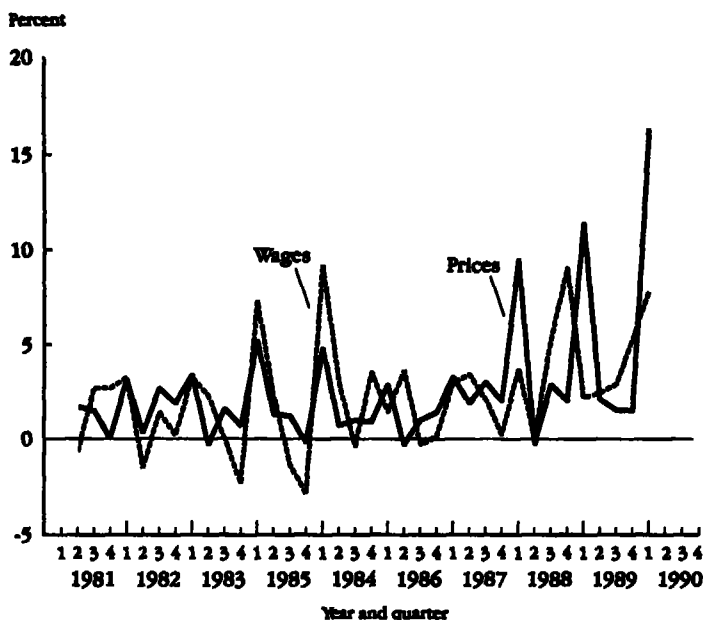
Figure 2. *Changes in Prices and the Official Exchange Rate in Poland, 1982.2–89.4*

Source: International Monetary Fund (various years).

Hungarian inflation, particularly during the pre-1982 period.[1] There is no evidence of structural change in Hungary up to 1987. Inflation has had a marked inertial component, as would be expected in a system with a significant share of administered prices. There is strong covariation of cost factors and domestic prices. Changes in unit labor costs and excess demand appear to have been more powerful determinants than first-round import price effects of the inflation rate.

Inflation in Poland. In Poland during 1953–82, when prices were fully controlled from the center, consumer prices broadly reflected movements in total costs. However, within this aggregate movement, large changes in relative prices took place. Prices of basic items—food, transport, rent, and energy—lagged consistently behind the overall price index. In addition monetary growth outpaced price growth in the 1970s, creating a significant imbalance, which was further associated with a large premium of the black market exchange rate over the official exchange rate.

The price reform of 1982, with the attendant jump in administered prices, aimed to change both relative prices and reduce the monetary overhang. The

1. Results of two-stage least-squares estimation of the annual inflation model for Hungary are available from the authors.

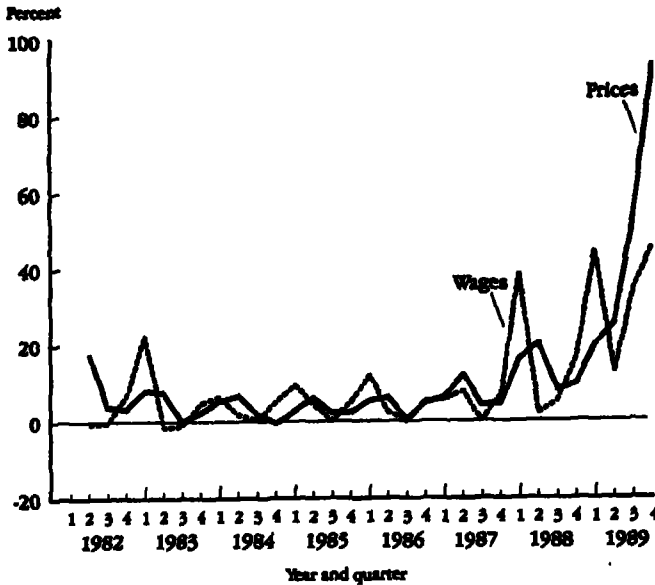
Figure 3. *Changes in Prices and Wages in Hungary, 1981.2–90.1*

Source: International Monetary Fund (various years).

initial jump in the price level translated into an upward shift of the ra
inflation, which stabilized before 1987 at around 15–20 percent a year. En
cal evidence, including tests for structural breaks, suggests that the period
1982 represents a new regime in price behavior and not simply an increase i
rate of inflation.[2] This result seems consistent with the qualitative shift fr
system of fully controlled prices to a mixed system of controlled and, in pı
ple, market-based prices.

The 1980s were marked by major external imbalances, persistent cu
account deficits, and pressure on the exchange rate. The black market exch
rate premium remained very substantial throughout the 1980s—one indicat
the presence of repressed inflation in the system. Before 1989, howeveı
frequent devaluations of the official exchange rate did not succeed in redı
the size of the premium.

An important factor for explaining the jump into high inflation in 19
Poland was a sharp deterioration in the fiscal accounts. Since 1981 the
deficit had generally been restrained to under 1 percent of gross national prc
(GNP), and in 1988 it was held to 0.22 percent. In the first half of 1989 rev
losses amounting to more than 7 percent of GNP pushed the fiscal deficit clo

2. Chow tests and recursive estimation results are available from the authors.

Figure 4. *Changes in Prices and Wages in Poland, 1982.2–89.4*

Source: International Monetary Fund (various years).

percent of GNP. Continuous depreciation of the exchange rate, liberalization of food prices, major shocks to administered prices, and the introduction of an ex-post indexing scheme in July 1989 generated a wage-price spiral that tipped the economy into hyperinflation. Money could not anchor the system in the absence of exchange rate and wage stability. Further, with the lifting of price controls and the large increases in administered prices, neither could the price controls anchor the system adequately. Sharp upward adjustments to administered input prices in late 1989 were initially passed through to final prices by producers.

The deceleration of inflation after February 1990 can be attributed not only to a major fiscal correction but to the reinstallation of anchors in the system (Coricelli and Rocha 1991). The exchange rate and wages have been the principal anchors, but temporary restraint on adjustments to remaining administered prices has also been applied. Measures to break the accommodating role of money have been implemented through the independence of the central bank, interest rate policy, and credit ceilings. Tight money, input price adjustments, and a fall in domestic effective demand, alongside a trade liberalization aimed at importing international prices and a new relative price structure, led producers in "exposed" sectors—particularly consumer goods—to lower markups.

Markups for producer goods were sustained longer (Frydman and Wellisz 1991). These measures reduced monthly inflation to less than 5 percent by October 1990. However, output declined by around 28 percent in the industrial sector during the first eight months of 1990, and unemployment increased to 7.5 percent of the labor force.

Wage Determination

Wage determination is critical as there is no effective endogenous mechanism of wage restraint in a socialist economy. In Hungary centrally imposed wage norms and bargaining rules were largely sustained, and wage drift was restricted. For Poland wage bargaining can best be interpreted in a noncooperative setting (Commander and Staehr 1990). The outcome was stronger wage and cost-side pressure on prices. Polish inflation was consequently powerfully linked to changes in the two key flow variables: wages and the exchange rate (Lipton and Sachs 1990).

II. The Model

The model summarized here describes a simple markup pricing system in which price movements, either through government controls or through the behavior of monopolistic firms, reflect cost movements (for detailed features, see the appendix). The pricing routine is related to the absence of any strong, market-based restraint on either the price or wage side. Consequently, both price and wage controls are intrinsic to the stability of the system. Prices and market power are restrained by permissible margins on the markup, and wages are restrained by reference to expected inflation and some association to productivity.

Features of the Partially Reformed Socialist Economy

The sustaining conditions for price growth can be traced both to the fiscal-monetary side and to the discrete effects of exchange rate adjustment. On the fiscal-monetary side, the rate of monetization will be directly acted on by the consolidated fiscal deficit, including the quasi-fiscal deficit. Sustained inflation can be attributed to the presence of fiscal deficits that cannot be financed through debt issue on a sustainable basis (Anand and van Wijnbergen 1987).

In systems in which administered prices are significant, if price controls repress inflation in the short run, the elimination of budgetary subsidies associated with lifting price controls would translate that repressed inflation into open inflation. Moreover, with regard to the exogenous shift factors affecting the inflation rate, specific weight has to be attached to the role of discrete adjustments to administered prices by the planner. Such adjustments can be viewed as a lagged response by the planner to macroeconomic imbalances, particularly excess demand in goods markets, and as an attempt to modify the structure of relative prices. However, excess demand would be determined not simply by

anticipated changes in foreign price, monetary, and fiscal impulses, but also the effect of fixed-price regimes.

Dynamic adjustment is also affected by the timing of price decisions. Administrative rules generally impose long lags between adjustments of controlled prices. A mixed system of controlled and market prices leads to two sources of nonsynchronization in the price setting (Commander and Coricelli 1990). The first relates to the nonsynchronization between the change of controlled and market prices. The second arises, as in market economies, from the decentralization, and thus lack of coordination, of the market.

Liberalization of prices for certain categories of goods is likely to increase the flexibility of the price system. However, liberalization introduces an element of nonsynchronization in price changes and thus can increase price system inertia (Taylor 1979; Pazos 1972). Both price staggering and wage indexation could be expected to facilitate the translation of shocks to the price level into a persistent increase in the rate of inflation. The persistence of inflation is strengthened to the extent that enterprises operate a constant markup rule, wages are inflexible, and the planner adjusts administered prices in lagged response to excess demand in goods markets.

Reduced-Form Price and Wage Equations

The reduced-form dynamic models for prices and wages are given in equations 1 and 2:

$$(1)\quad Dp_t = \alpha_0 + \alpha_1 Dp_{t-1} + \alpha_2(Dpw_t + De_t) + \alpha_3[Dw_t - D(y/L)_t] + \alpha_4(y_t - y^*) + \alpha_5\,[(M/RS)_{t-1} - (M/RS)^*_{t-1}] - \alpha_6(p - p^*)_{t-1}$$

$$(2)\quad Dw_t = \beta_0 + \beta_1 Dw_{c_t} + \beta_2[D(y/L)_t] + \beta_3(w - w^*)_{t-1} + \beta_4 De_{B_t} + \beta_5[(RS/P)_t - (RS/P)^*_t].$$

In equation 1 prices (p) are a function of lagged prices, import prices (e denotes the exchange rate), unit labor cost (w, y, and L denote wages, output, and labor input, respectively), capacity utilization, an excess demand variable (constructed as the deviation from a perceived normal ratio of monetary holdings in the household sector [M] to retail sales [RS]), and a correction term. In equation 2 wages are related to the centralized wage norm (w_{c_t}), productivity, a correction term embodying a real wage target (w^*), a black market exchange rate premium variable (e_B) for Poland, and a variable that attempts to capture the sensitivity of wage demands to the availability of goods in consumer markets (the deviation from trend of real retail sales for consumer goods).

The specification of the correction terms implies convergence to equilibrium values or to certain target values ($p - p^*$ in equation 1 and $w - w^*$ in equation 2). However, because socialist economies are commonly thought to be disequilibrium regimes, two aspects have to be emphasized. First, an error correction implies the presence of feedbacks, which induce changes in the dependent va-

able when the system deviates from its long-run point of attraction. The feedbacks are total costs or money for prices, and prices or productivity for wages. This does not imply any concept of equilibrium, such as market clearing, because such systems lack endogenous equilibrating mechanisms operating through the labor market or capacity. However, underlying regularities or rules of price and wage behavior were fairly stable during the 1980s. Interestingly, for pricing, these rules were embedded in the institutional setting in which price changes explained by cost changes were permitted and justified. The error correction approach implies that the dynamics are affected by a tendency toward these underlying rules. Monetary variables are not included in the correction term, despite the fact that domestic money and prices are cointegrated, because money is passive in such economies.

III. Short-Run Price and Wage Dynamics: Results from Quarterly Data

The reduced-form equations are estimated using quarterly data from Hungary and Poland. The specified periods are in part given by data availability. To explore the price-wage interdependency and to avoid simultaneity bias, we employ the three-stage least-squares procedure.[3]

Hungary: 1981.3—89.2

The results of simultaneous estimation of the price and wage equations using quarterly data for Hungary are reported in table 2. Consumer price changes are related to lagged prices, the exchange rate, contemporaneous and lagged unit labor cost summarized in a first-period moving average, a money term proxying demand-side features, and a correction term relating prices to unit labor costs. Wages are related to consumer prices, productivity in the material sector, and a correction term, specifying a real wage target.

The variables are stationary and significant at the 5 percent level. For the price equation the overall fit is quite acceptable, there is parameter stability when estimated recursively, and there is an absence of serious autocorrelation. The equation is not sensitive to altering the sample period. The wage equation is more problematic, appearing underidentified and with its fit deteriorating seriously toward the end of the sample. Nevertheless both equations capture the short-run dynamics reasonably well.

The estimations indicate some key dynamic features for price and wage changes during the 1980s. Although foreign prices are important in the trans-

3. Before estimating the dynamic models, we tested for stationarity with Dickey Fuller and Augmented Dickey Fuller tests. We recognize that these are weak tests, but conveniently the time series proved to be largely integrated of the first order, I (1), so that first differencing of the data was appropriate. Estimation was initially done with levels and then in dynamic form with truncation of the lag polynomials and some restrictions on coefficients. The correction terms represent the lagged residuals from discrete, truncated levels equations, tested for nonstationarity of the residuals. Finally, the equations were tested for parameter stability and for sensitivity to changes in sample period.

[Tabl]e 2. *Estimation Results from Quarterly System Three-Stage Least Squares, [Hun]gary 1981.3–89.2*

Item	*Estimated value*
Dependent variable: consumer prices	
Consumer prices lagged one period	0.23 (2.70)
Official exchange rate	0.23 (4.11)
Unit labor cost[a]	0.65 (6.49)
M2/retail sales	0.21 (2.22)
Correction term[b]	−0.14 (−1.87)
$R^2 = 0.80$	
se = 0.016	
F = 17.73	
DW = 1.87	
Dependent variable: wages	
Consumer prices	0.86 (5.94)
Industrial productivity[c]	0.19 (1.89)
Dummy variable for 1984.4	−0.05 (−2.32)
Dummy variable for 1988.1	−0.05 (−2.26)
Correction term[d]	−0.52 (−3.65)
$R^2 = 0.65$	
se = 0.022	
F = 9.91	
DW = 1.82	

Note: All variables except for the dummy variables are in first differences of logs; *t*-statistics are in parentheses. Cointegration tests and results are available from the authors.

a. One-period moving average of first differences of the log of wages minus the log of productivity.

b. Correction term is lCPI $= c_1 + c_2$ lULC (ULC is unit labor cost) lagged one period.

c. Industrial production/industrial employment.

d. Correction term is lWAGE $= c_1 + c_2$ lCPI (CPI is consumer price index) lagged one period.

Source: International Monetary Fund and authors' estimations.

mission of inflation, cost-side factors have greater weight. Demand-side pressures matter, and it seems likely that one mechanism works through the impact on the relative prices of administered and market goods as a factor determining the scale of adjustment to administered prices. Strikingly, no standard association between output or capacity utilization and prices exists. This emphasizes the lack of an underlying mechanism of equilibration in the economy despite the introduction of greater market-based features. The presence of administered prices imparts inertia to the system, but the models, as presently specified, cannot explicitly capture the price-price dynamics associated with nonsynchronization in price adjustment over administered and market prices.

The price equation. A lagged price term is included on the right side. Argument for inclusion of this term relates to the weight of administered prices where such prices can be considered to have stronger-than-normal properties of exogeneity. The coefficient on a one-period lag at 0.23 points to the fairly high degree of inertia in the system. The impact elasticity with respect to the exchange rate at 0.23 is greater than in the annual model, but this difference could be expected given the more active exchange rate policy and stronger explicit linkage of domestic to foreign prices during this period. The demand pressure variable is very significant. Little can be said about the interactive effects on administered and market prices of demand pressure, but the elasticity indicates the somewhat diluted translation into aggregate price increase.

In the initial single equation estimation, we inserted an output gap variable. It could have been expected that with the progressive introduction of market-based rules into the Hungarian economy that a more conventional relation of output to prices would emerge. Recursive estimation and shortening of the sample at the end of the period did indeed result in the term's picking up significance toward the end quarters. However, closer examination of the residuals indicated that the significance was largely taken from the noise associated with the price shifts during and after 1988, and the term was not included in the final estimations. The unit labor cost variable, with a moving average elasticity of 0.65, substantiates the powerful link of price growth to costs, as would be expected from the model elaborated above. The correction term is the argument that prices are cointegrated with unit labor costs. The Dickey Fuller test exceeds the 5 percent critical value for the null hypothesis for this specification. The coefficient enters the equation with the right sign and is significant at the 5 percent level. The mean lag of adjustment is reasonably rapid at around four quarters.

Recursive estimation and Chow tests prompt inclusion of a dummy for the first quarter of 1988. There is no serious autocorrelation, even when the lagged-price term is dropped. The bands for the recursive estimations widen slightly at the end of the period, pointing to larger unexplained variation at the end of the sample. One-step residuals confirm this feature while Chow tests show a lack of smoothness toward the end of the period, with a persistent spike at 1988.1. This is unsurprising given the timing of changes to administered prices and the tax system in this period. The dummy for 1988.1 in the single equation estimation was significant but did not change the size of the other coefficients or improve the overall fit or properties of the model. Consequently, in the final system estimation it was omitted.

The wage equation. The wage equation appears underidentified, and there is some negative autocorrelation in the residuals. There is a very marked deterioration in performance at the end of the sample. No wedge term could be adequately specified capturing any gap between producer and consumer wage, and there is no explicit treatment, due to lack of data availability, of the wage tax

Up to 1985 wage changes were clustered in the first quarter; thereafter [an]nual wage round gave way to more random adjustments. Wage adjust[ments] within years appear to have been adopted as an institutional routine [associ]ated with the acceleration in consumer prices during the later part of the [decad]e. For consumer and producer prices there is a similar clustering of [chan]ges in the first quarter. For productivity—measured over the socialist mate[rial s]ector—changes generally occur in second and fourth quarters. The spike in [the] fourth quarter can partly be explained by the structure of the wage round [and] the effect of the wage tax in associating wage and productivity paths. [Sta]bility tests picked out breaks at 1984.4 and 1988.1, for which dummies were [ins]erted. In general, institutional particularities, changes to the wage system, [an]d wage tax rules and their incomplete specification in the model explain the [lim]itations of the reported equation (Commander and Staehr 1990).

The 0.86 coefficient on the price term in the wage equation points to the [st]rong association of wage to consumer price movements. There appear to be no [ef]fective unanticipated price changes or price surprises. Testing this further by [in]serting a price surprise variable yielded a weak positive coefficient that was [v]ery insignificant. The absence of any price surprise, despite the acceleration in [in]flation during this period, can best be explained with respect to a real wage [ta]rget maintained by the planner. Wages appear to have been fixed in relation to [a]nnounced prices. To the extent that centrally given wage norms have aimed to [im]pose an inflation penalty, bargaining has reconverged the wage and price [p]ath.

The productivity term is significant, and its coefficient has roughly the same [m]agnitude as that for market economies. The correction term is written as a real [w]age target, is very significant, and points to a rapid adjustment speed. It should [h]owever be interpreted with some caution as an observed association between [w]age and price variables during the sample period, rather than as a long run [e]quilibrium relationship. It is evident that there is an underlying linkage between [p]rice, productivity, and wage changes, as suggested by our model. Conse[q]uently, the wage does not enter as a completely exogenous variable.

Poland: 1983.4—90.1

The Polish price system has been subjected to a set of major institutional [c]hanges. The first change was the creation of a mixed system of free and con[t]rolled prices in 1982. There were further modifications in 1988–89 and a major [c]hange in January 1990. Using annual data, Chow tests and recursive estima[t]ion identify 1982 as a structural break. Our estimations consequently use quar[t]erly information for the period after 1982. Preliminary testing for a break in [1]990.1 does not indicate a comparable break in the pricing regime.

The estimation results for Poland are presented in table 3. Before estimating [th]e system, price and wage equations were run discretely to test for parameter [s]tability, data coherence, and goodness of fit. All variables are highly significant. [T]he model, as shown by stability tests, survives through the period of high,

Table 3. *Estimation Results from Quarterly System Three-Stage Least Squares, Poland 1983.4–90.1*

Item	*Estimated value*
Dependent variable: consumer prices	
Consumer prices lagged one period	0.33 (4.20)
Official exchange rate	0.33 (7.60)
Unit labor cost[a]	0.43 (5.90)
M3/retail sales	0.34 (5.40)
Correction term[b]	−0.17 (−1.60)
$R^2 = 0.98$ se = 0.04 F = 220.02 DW = 1.88	
Dependent variable: wages	
Consumer prices	0.68 (16.40)
Industrial productivity	0.58 (3.10)
Deviation from trend of real retail sales	0.20 (2.80)
Dummy variable for 1988.1	0.20 (5.80)
Dummy variable for 1989.1	0.24 (6.60)
Correction term[c]	−0.40 (−4.50)
$R^2 = 0.93$ se = 0.056 F = 60.86 DW = 2.12	

Note: All variables except for the dummy variables are in first differences of logs; *t*-statistics are in parentheses. Cointegration tests and results are available from the authors.

a. One-period moving average of first differences of the log of wages minus the log of productivity.

b. Correction term is $\text{lCPI} = c_1 + c_2 \text{lULC} + c_3 \text{lOER}$ (ULC is unit labor costs, and OER is official exchange rate) lagged one period.

c. Correction term is $\text{lWAGE} = c_4 + c_5\text{lCPI} + c_6\text{lPRO}$ (CPI is consumer price index, and PRO is productivity in the industrial sector) lagged one period.

Source: International Monetary Fund and authors' estimations.

almost hyper, inflation of 1989–90. This seems to suggest that administered price changes, staggered adjustment of other prices, wage indexation, and an effectively indexed exchange rate maintained, if not increased, inertia in the inflation process.

For the price equation the fit is very acceptable, and the variables are stable. For the wage equation the fit is very satisfactory, but there is clear evidence (detected through recursive estimations and Chow tests) of a large spike at 1988.1 and 1989.1. Dummy variables are included for both periods. These

mmies can be interpreted as representing some independent wage push, which :urred in a period of increasing social unrest.

The price equation. Prices were regressed on a lagged price term, a cost riable, the official exchange rate, a lagged demand pressure variable, and a rrection term. A lagged price term is included for the reasons already alluded in the Hungarian model. In this specification the measure of money was M3, :luding foreign currency deposits, deflated by retail sales. Prices are cointe- ıted with unit labor costs and the official exchange rate.

A cost-plus specification is used, linking the short-term dynamics to an adjust- nt to an underlying normal markup pricing rule. This specification includes netary variables only as determinants of short-run movements, and not of a ıg-run equilibrium relation. Empirical tests were run for the presence of a nificant long-run relation between prices and money. A simple standard ney demand approach indicated the presence of cointegration between prices 1 money. However, the inflation transmission mechanism is very poorly re- :ted by a dynamic equation based on this long-run tendency. This strengthens view that the concurrent movement of prices, wages, and monetary aggre- es in the long run is more the result of monetary accommodation than of a ısal relation running from money to prices.

nflation is characterized by a relatively high degree of inertia throughout the iod. This is captured by the presence of a lagged inflation term in the regres- n and by a relatively low coefficient of the correction term, implying a some- at slow reaction to deviations from a normal relation between prices and ıts. This pattern accords well with the observation of a relatively high and ble inflation during 1984–87. Moreover the stability of the relation through- : 1983–90 suggests that this degree of inertia survived during the acceleration inflation after 1987. Although the coordination problem typical of systems h many price setters did not affect Poland, the inertia in inflation may have sen from the peculiar features of a mixed system of market and controlled :es, in addition to the high degree of de facto indexation of wages to prices ›mmander and Coricelli 1990). Different rules and timing of price decisions the different price categories likely gave rise to a substantial staggering of :e decisions and a consequent inertia in inflation.

'he wage equation. In the wage equation, wages were related to productivity :he material sector, prices, a variable denoting deviation from trend of real ıil sales, a correction term, and dummy variables for 1988.1 and 1989.1. e correction term reflects cointegration between wages, prices, and ductivity.

The wage equation helps to stress the role of wages in the transmission of ationary impulses, especially during 1988–90. Wages have been de facto ıost fully indexed to prices throughout the period. Besides the inflation term, other variables entering the equation show an effect of productivity on

wages, an effect of availability of goods on wage demands, and some effect of changes in the premium of black market over official exchange rate on wage movements. The latter effect may indicate that workers are concerned with the purchasing power of wages, not only in terms of domestic goods but also of dollar goods. This may be explained by the presence of shops selling foreign goods for foreign currency or by the use of the premium as a proxy for prices in nonofficial markets.

The relation between wages and productivity is somewhat puzzling. The rather high coefficient may be partly explained by the system of bonus payments and by a similar seasonal pattern of wages and productivity changes. The high (0.20) elasticity of wages with regard to the demand pressure variable (deviation from trend of retail sales) is also striking. Contrary to theories of wage illusion, which might have been relevant in the 1960s and 1970s (Charemza and Gronicki 1988), the perception of a greater availability of goods in the market may have exerted an upward pressure on wage demands. The correction term can be interpreted as a real wage target, adjusted for productivity changes and the resulting payment of bonuses. Moreover the speed of adjustment to deviations of wages from the target level is rather rapid. The coefficient of the correction term, together with a coefficient of nearly 0.7 on the price term, points to a high degree of rigidity of real wages, both in the changes and in the levels.

IV. Concluding Remarks

This article sets up a simple inflation model centered on the transmission process and on the short-run dynamics of inflation in partially reformed socialist economies. The model has features derived from market economies with few producers and sticky prices. It attempts to capture some of the specific attributes of socialist economies, including chronic excess demand in goods markets. The bulk of the empirical analysis concentrates on the period after 1982, when significant market-related reforms were implemented. Although socialist economies have been characterized by major disequilibria, we believe that appropriately specified and tested econometric models can account for these features, without explicit recourse to disequilibrium econometric procedures.

The dynamic price and wage models are estimated simultaneously. For both Hungary and Poland the estimation results allow us to explore the role and weight of foreign prices and domestic factors in propagating inflation. Foreign prices are shown to matter, but developments on the cost side are critical in relating exogenous, policy-determined adjustments to the price level to increases in the rate of inflation. Wages are indexed to prices for most periods; however, in Poland there are complex bargaining games and a corresponding inability to make centralized wage norms hold. Polish planners relied increasingly on price adjustments to address the macroeconomic imbalances that emerged, but these only further destabilized the system and failed to address the underlying sources of these imbalances. In contrast the Hungarian experience points to some of the

ays in which administered prices can be used to stabilize the system (Commander and Coricelli 1990).

An important implication is that nominal anchors, particularly wage restraints, should be included in stabilization programs in reforming socialist economies. The absence of conventional equilibrating mechanisms, and hence effective market restraints on wages alongside the presence of monopolistic pricing behavior, can easily result in powerful inflationary pressures and ultimately higher-than-warranted output costs.

Appendix. Details of the Price and Wage Models

The economy has two sectors: material and nonmaterial. Inflation is a weighted average of prices in the material sector (p_m) and prices in the nonmaterial sector (p_s):

(A-1) $$p = (1 - \alpha)p_m + \alpha p_s.$$

Each sector has a controlled and noncontrolled price subsector, denoted by subscripts c and n, respectively:

(A-2) $$p_m = \beta_1 p_{m_c} + \beta_2 p_{m_n}$$

(A-3) $$p_n = \theta_1 p_{s_c} + \theta_2 p_{s_n}.$$

For the material sector exposed to competition from the hard currency zone, export prices (p_x) are based on world prices (p_w) and the exchange rate (e):

(A-4) $$p_x = p_w + e.$$

Import prices are formed by world prices, the exchange rate, and tariffs (T):

(A-5) $$p_m = \alpha T(p_w + e).$$

In equation A-5 α denotes a terms-of-trade parameter. Total prices in the material sector trading outside the system of the Council for Mutual Economic Assistance would be a weighted average (k) positively associated with the share of controlled to total prices. Thus

(A-6) $$p_m = k(p_w + e).$$

The following markup pricing rule is assumed, where p_o denotes output prices in the nonmaterial sector and that part of the material sector isolated from the competition of hard-currency goods:

(A-7) $$p_o = (1 + \mu) \frac{p_i I + wL}{y}$$

where μ is the rate of markup, p_i is the price of inputs, I is inputs volume, w is wages, L is labor volume, and y is output.

The markup rule has standard properties. Changes in output prices will thus depend on unit factor costs and on the markup. Foreign prices do not matter. To

the extent that the markup relates to some normal level of output (Coutts, Godley, and Nordhaus 1978), the lower the weighted average in equation A-6, the more likely it is that price change will be positively associated with the level of capacity utilization. Capacity utilization (cu) can be written in the usual way as deviation of output from trend (y^*):

$$cu = (y - y^*). \tag{A-8}$$

The greater the weight of controlled prices in total prices, the weaker the link between price changes and capacity levels. One would therefore expect a classical planned economy to lack any conventional equilibrating mechanism. In addition this introduces a relative price effect so that the coefficients depend not only on components of costs, but also on the elasticity of the markup with respect to the overall domestic price index. In terms of rates of change, this yields the following linearized, preliminary price equation:

$$p_t = \alpha_0 + \alpha_1 p_{t-1} + \alpha_2(p_w + e) + \alpha_{3a}(wL/y) + \alpha_{3b}(p_i I/y) + \alpha_4(y - y^*). \tag{A-9}$$

Prices are a positive function of the markup, foreign prices (through α_2 and α_3), deviations from trend of capacity utilization, and lagged prices. Price setting in the nonmaterial sector would be a subset of equation A-9, obtained by omitting α_2.

Equation A-9 closely resembles a standard markup model applied extensively to nonsocialist economies. We now accommodate some stylized features of socialist economies: the widespread presence of goods shortages and high liquidity ratios. Purchasing power in consumer markets may generate undesirable effects in the view of the planner. The presumed concern with relative prices on the part of the planner suggests that demand pressure in goods markets can best be captured in a price equation by a variable $[(M/RS) - (M/RS)^*]$ measuring deviation from a perceived normal ratio of monetary holdings (M) in the household sector to retail sales (RS). This indicates the measure of excess purchasing power in consumer markets and hence the assumed demand-side pressure for relative price adjustment by the planner.

The inclusion of both an output gap ($y - y^*$ in equations A-8 and A-9) and an explicit purchasing power variable is an attempt to pin down the particular excess demand features of these economies when controlled and market prices coexist and when the planner monitors their relative price. To the extent that supply remains insensitive to price changes, one would expect a strongly damped equilibrating relation between shortage and inflation.

We can now set up in dynamic form the following inflation equation, where D is a first difference operator and α_6 is the correction term.

$$Dp_t = \alpha_0 + \alpha_1 Dp_{t-1} + \alpha_2(Dp_{w_t} + De_t) + \alpha_3[Dw_t - D(y/L)_t] + \alpha_4 D(y_t - y^*) + \alpha_5[(M/RS)_{t-1} - (M/RS)^*_{t-1}] - \alpha_6(p - p^*)_{t-1}. \tag{A-10}$$

Wages

An excess demand for labor regime obtains in the socialist economy. No usual ;onvex relationship between labor market pressure and wage growth exists. Restrictions on profit-maximizing behavior at the level of the enterprise imply, n combination with full employment, that the wage share would not necessarily se endogenously restrained by productivity and prices. Wages are treated largely is exogenous and structured by centrally determined norms. Productivity-related wage components—a feature of all partially reformed economies—ntroduce a crude association to average productivity, rather than marginal products.

Distinguishing again between material and nonmaterial sectors, aggregate rage growth is a weighted average. In so far as total wages converge to roductivity-corrected wage levels for the material sector, a faster rate of pro-uctivity increase in the material sector would, other things being equal, acceler-te inflation.

Wage contracts are of annual duration, with explicit commitment to real wage .ability. Available evidence indicates in general adherence to a real wage target Adam 1979). In principle a form of forward-looking or ex-ante indexation iould be operating. Indeed increases related to the centrally imposed norm :flect expected or planned inflation. In this case

$$Dw_c = \alpha_1 + \alpha_2 E_{t-1}(Dp) \quad \text{(A-11)}$$

'here w_c is the centrally imposed wage, E is the expectation operator, and p is rice inflation. The real wage is a decreasing function of expectational errors, or iflation surprises:

$$Dw_c = \alpha_1 + \alpha_2 Dp - \alpha_3[Dp - E_{t-1}(Dp)]. \quad \text{(A-12)}$$

n ex-post indexation scheme would imply

$$Dw_c = \alpha_1 + \alpha_2 Dp_{t-1}. \quad \text{(A-13)}$$

Any increase in the inflation rate reduces the real wage, while the indexation :heme introduces inertia in the inflationary process, slows down a reduction in ie inflation rate, and makes the wage setting insensitive to any announcement 'fects (which in the ex-ante specification will immediately reduce the increase in 'ages). Combining these features with the postulated relation between wages id average productivity yields the following equation to be estimated:

$$Dw_t = \beta_0 + \beta_1 Dw_c + \beta_2[Dq_m + Dp - D(1 + \tau)] \quad \text{(A-14)}$$

here q_m is average productivity in the material sector and τ is the tax rate on age growth exceeding the centrally imposed wage, adjusted for productivity. 'age growth consequently derives from a centralized component related to pected inflation and drift attributable to productivity changes in the material ctor, offset by tax increases.

Equation A-14 would not adequately capture the persistent increase in real ages up to the 1970s and apparent episodes of wage push during the 1980s in

Poland. Assuming bargaining and a concern of workers with a real wage target, the insertion of an error correction term in the equation appears appropriate:

$$\text{(A-15)}\quad Dw_t = \beta_0 + \beta_1 Dw_c + \beta_2[Dq_m + Dp - D(1+\tau)] + \beta_3(w - w^*)_{t-1}$$

where the wage target (w^*) is equal to price inflation.

As with the price equation, equation A-15 is modified by introducing features specific to a socialist economy. Two variables may affect the wage dynamics. The first is an actual or implicit price index, which incorporates unofficial markets. The premium of the black market exchange rate over the official exchange rate (e_B) is used as a proxy for this unrecorded inflation. In economies characterized by chronic shortages, availability of goods may significantly affect wage demands. Thus the second variable specific to a socialist economy measures the deviation from trend of inventories of consumer goods at the retail level. The second variable is denoted by $(RS/p)t - (RS/p)_t^*$, where RS is retail sales.

The general form for the dynamic wage equation now becomes

$$\text{(A-16)}\quad Dw_t = \beta_0 + \beta_1 Dw_{c_t} + \beta_2[Dq_{m_t} + Dp_t - D(1+\tau)] + \beta_3(w - w^*)_{t-1} + \beta_4 De_{B_t} + \beta_5[(RS/p)_t - (RS/p)_t^*].$$

However, as we maintain that consumer goods market imbalances and currency substitution have been more powerful in the Polish case, we will drop these terms at the outset from the Hungarian estimations.

References

The word "processed" describes informally produced works that may not be commonly available through library systems.

Adam, Jan. 1980. *Wage Control and Inflation in the Soviet Bloc Countries*. New York: Praeger.

Anand, Ritu, and Sweder van Wijnbergen. 1987. "Inflation and the Financing of Government Expenditure in Turkey: An Introductory Analysis." World Bank, Washington, D.C. Processed.

Charemza, Wojciech, and Mirosaw Gronicki. 1988. *Plans and Disequilibria in Centrally Planned Economies: Empirical Investigation for Poland*. Amsterdam: North-Holland.

Charemza, Wojciech, and R. E. Quandt. 1982. "Models and Estimation of Disequilibrium for Centrally Planned Economies." *Review of Economic Studies* 49 (155) :109–16.

Commander, Simon, and Fabrizio Coricelli. 1990. "The Macroeconomics of Price Reform in Socialist Countries: A Dynamic Framework." PRE Working Paper 555. World Bank, Economic Development Institute, Washington, D.C. Processed.

Commander, Simon, and Karsten Staehr. 1990. "The Determination of Wages in Socialist Economies: Some Microfoundations." WPSEDI 713. World Bank, Economic Development Institute, Washington, D.C. Processed.

ɪricelli, Fabrizio, and Roberto Rocha. 1991. "Stabilization Programs in Eastern Europe: A Comparative Analysis of the Polish and Yugoslav Programs of 1990." In Vittorio Corbo, Fabrizio Coricelli, and Jan Bossak, eds., *Reforming Central and Eastern European Economies: Initial Results and Challenges*. Washington, D.C.: World Bank.

ɪutts, Kenneth, Wynne Godley, and William Nordhaus. 1978. *Industrial Pricing in the United Kingdom*. Cambridge University Press.

'tenstein, Andrew. 1989. "Inflation, Price Controls and Fiscal Adjustment: A Survey of the Literature and Analytical Framework." University of Kansas Lawrence. Processed.

/dman, Roman, and Stanislaw Wellisz. 1991. "The Ownership-Control Structure and the Behavior of Polish Enterprises during the 1990 Reforms." In Vittorio Corbo, Fabrizio Coricelli, and Jan Bossak, eds., *Reforming Central and Eastern European Economies: Initial Results and Challenges*. Washington D.C.: World Bank.

ɪnohan, Patrick. 1989. "Inflation in China." World Bank, Washington, D.C. Processed.

ernational Monetary Fund. Various years. *International Financial Statistics*. Washington, D.C.

ɪton, David, and Jeffrey Sachs. 1990. "Creating a Market Economy in Eastern Europe: The Case of Poland." *Brookings Papers on Economic Activity*. 1:75–147.

ti, D. M. 1986. "Hidden and Repressed Inflation in Soviet-Type Economies: Definitions, Measurements and Stabilization." In *Contributions to Political Economy*. Published for the Cambridge Political Economy Society. London: Academic Press.

:os, Felipe. 1972. *Chronic Inflation in Latin America*. New York: Praeger.

'tes, Richard. 1977. "The Control of Inflation: Lessons from East European Experience. *Economica* 44 (174): 109–30.

cha, Roberto. 1990. "Structural Adjustment and Inflation in Yugoslavia." World Bank, Washington, D.C. Processed.

'lor, John. 1979. "Staggered Wage Setting in a Macro Model." *American Economic Review Papers and Proceedings*.

WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 55-69

Are High Interest Rates Effective for Stopping High Inflation? Some Skeptical Notes

Guillermo A. Calvo

High interest rates are considered an effective tool for stopping high inflation. The case for a policy of high interest rates is developed in terms of a conventional IS-LM model. However, among other things, the model ignores some central aspects of modern credit markets. In particular high interest rates may give rise to nonperforming bank loans, thus seriously jeopardizing the effectiveness of the policy. Examples are developed in which it would be optimal to aim for equilibriums of low, rather than high, interest rates. One of these examples hinges on the existence of nonindexed domestic debt.

This article looks at the effectiveness and desirability of relatively high interest rates in stabilization programs. High interest rates refer to nominal interest rates that would result in unsustainable, high real interest rates—that is, real rates of interest that would cause a state of generalized bankruptcy if maintained for a few months or a year—if the program's inflation target were to be achieved.

The article evolves through a series of examples, beginning with the *IS-LM* model described in section I. Despite its simplicity (and early vintage), the model is still important in discussions of policy. The conventional *IS-LM* approach, which gives strong support to a policy of high interest rates, implies that tight money is effective in reducing economic activity and, through the Phillips curve, is also effective in lowering inflation. The larger the interest rate, the sooner price stability will be achieved.

Section II questions the relevance of the *IS-LM* result on the basis of two observations. First, we live in a world in which countries are closely linked from a financial point of view. Therefore high interest rates are likely to signal expectations of high devaluation or inflation. Second, high inflation rates make expected or ex ante real interest rates look much smaller than if the stabilization plan's inflation targets were to be fulfilled. Consequently, banks may continue lending despite high nominal rates. With the passage of time, however, firms will find themselves in serious financial straits, which eventually may result in a state

Guillermo A. Calvo is in the Research Department at the International Monetary Fund. The author wishes to thank Bijan Aghevli, Morris Goldstein, Tim Lane, and three anonymous referees for their useful comments.

of generalized bankruptcy. This phenomenon may grow almost unnoticed until it is too late to react within the limits of the original program.

Section III discusses "sustainability" issues by studying the long-run implications of raising the interest rate on bank deposits. A high rate of interest on bank deposits provokes results that are similar to a cut in the money supply, but does not call for an overall credit rollback. Thus the high rate may look attractive in the short run. However, in contrast with the *IS-LM* approach, the longer-run perspective illustrated in section III shows that a 1 percent increase in the rate of interest on bank deposits translates immediately into a rate of inflation that is 1 percent higher.

Section IV makes the case for particular instances in which interest rates or related policy instruments may be used to eliminate equilibrium indeterminacy and, thus, facilitate the success of a stabilization program. Equilibrium indeterminacy may be a serious problem if there exists a sizable stock of nonindexed government debt and the government is perceived as behaving in a discretionary manner (that is, the government is unwilling or unable to commit itself to a predetermined set of policy rules). Debt indexation may be a way to avoid equilibrium multiplicity and coax the system toward the best equilibrium outcome.

Section V highlights the special circumstances of previously centrally planned economies undertaking radical economic transformation programs. It describes conditions under which a policy of high interest rates may be useful and the factors in Eastern European countries that would limit the usefulness of such a policy.

I. The Conventional Model

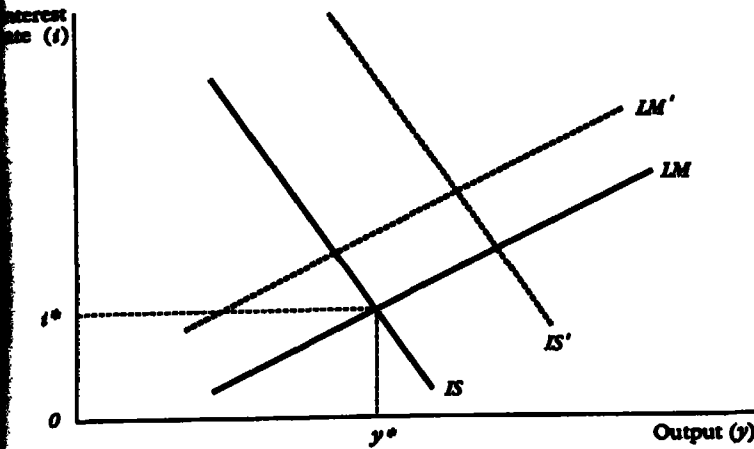
Much of current macroeconomic intuition owes a great deal to the enormously successful *IS-LM* model, originally developed by Hicks (1937). The model can be expressed by two market-clearing conditions, corresponding to the money and goods markets. The following log-linear equations represent closed-economy versions of the *IS-LM* model:

$$m - p = a - b(i - i_m) + y \text{ (money market)} \quad (1)$$

$$y = c - d(i - \pi^e) \text{ (goods market)} \quad (2)$$

where m, p, and y denote the (logs of) the money supply, price level, and output; while i and i_m stand for the nominal interest rate on loans and on money. The expected rate of inflation is denoted by π^e. Parameters a, b, c, and d are all positive.

The left side of equation 1 is the real supply of money; the right side is the demand for real monetary balances. Since money yields a return, the model assumes that the demand for money is a negative function of the difference between its opportunity cost, i, and its return, i_m. Furthermore the demand for

igure 1. ***Determination of Output and Interest Rate***

Note: The *LM* curve shifts to *LM'* when the nominal interest rate on money increases. The *IS* curve shifts to *IS'* when the expected rate of inflation increases.

noney is unit-elastic with respect to output (this assumption can be easily re-
axed). The left side of equation 2 is the demand-determined level of output; the
ight side is output demand, which is assumed to be a decreasing function of the
xpected real interest rate.

The relations described in equations 1 and 2 are depicted in figure 1 for a
iven price level, inflationary expectations, and interest rate on money. Equa-
ion 1 is satisfied along the locus of points labeled *LM*, while equation 2 holds
long the line denoted *IS*. The equilibrium levels of the nominal interest rate and
output are indicated by i^* and y^*, respectively.

What does figure 1 say about the use of interest rate policy for stabilization?
The conventional wisdom relies on a Phillips curve relationship, according to
which the rate of inflation, π, is an increasing function of excess demand in the
goods market and the expected rate of inflation. This is represented by

$$\pi = j(y^* - y_{fe}) + \pi^e \tag{3}$$

where y_{fe} is full employment output and j is a positive parameter. From equa-
tions 1, 2, and 3, a cut in the money supply shifts the *LM* curve to the left, raises
the interest rate, and puts downward pressure on output, which lowers the rate
of inflation.

The same results could be obtained by raising the interest rate on money,
which would generate excess money demand and shift the *LM* curve to the left.
Several South American countries have used the interest rate on money for
stabilization. Manipulation of the interest rate on money does not require call-

ing back credit lines, as would be the case if the money supply were cut. Because the money supply normally includes bank deposits and bank deposits are the counterpart of bank loans, lowering the money supply sets in motion a contraction of bank credit, which might cause productive firms and financial institutions to collapse.

A rise in the interest rate on money would increase the equilibrium interest rate, which means that some investment projects would probably be dropped. This contraction, however, is likely to be more orderly than the one provoked by the indiscriminate curtailment of bank credit that a fall in the money supply might cause. The interest rate on money is an instrument that achieves the goals of a cut in the money supply at less cost in the short run. In the medium or long run the costs may be reversed, but the short-run benefits of a policy based on the interest rate on money may prove irresistible when a stabilization program begins to founder.

High (nominal) interest rates could result from high inflationary expectations. For example, a rise in the expected rate of inflation would shift the *IS* curve in figure 1 to the right, resulting in higher interest, but also higher output. As indicated in equation 3, inflation rises because expected inflation and output have increased. This underscores the relevance of lowering inflationary expectations.

II. A Policy of High Interest Rates: Elements for a Critique

A policy of high interest rates and tight money may be an effective tool to aid a disinflation process. However, persistently high interest rates may indicate lack of confidence in the program and may lead to a state of generalized bankruptcy. Unfortunately, it may take a long while for it to become apparent that serious financial trouble exists.

A recent example of a policy of high interest rates is the Argentinean stabilization program of July 1989. The nominal rate of interest hovered around 15 percent a month from August to November, while the exchange rate was fixed against the dollar and inflation came down to about 5 percent a month. The program collapsed toward the end of November 1989, and high inflation resumed during the next two months.

The Expected Rate of Devaluation and Inflation

The conventional model is a very short-run depiction of the economy, in which financial linkages with the rest of the world are not explicitly considered. Some countries with high inflation face serious international debt problems, and their public sectors are subject to stringent borrowing constraints. The private sector in these countries, however, may not face similar constraints. Because of capital flight and unofficial foreign exchange, portfolio shifts between foreign and domestic assets may completely bypass the need to use foreign credit lines.

herefore, at the margin, the domestic economy may function as if it operated der perfect capital mobility.

Under perfect capital mobility, the domestic interest rate, i, is equal to the ternational nominal interest rate (the interest rate in terms of foreign curncy), I, plus the expected rate of devaluation of the domestic currency, ϵ^e, as llows:

$$i = I + \epsilon^e + k \quad)$$

here I and k are assumed to be exogenous.

Given fixed exchange rates and the expected rate of devaluation, any attempt affect the domestic interest rate through monetary policy will simply provoke mpensatory capital flows. This is one of the main messages of the Mundellming extension of the conventional model (Frenkel and Razin 1987). A cut in money supply will induce the public to sell their dollars for domestic interestaring assets. The domestic currency will then tend to appreciate, and the ntral bank will be forced to intervene to keep the parity. The central bank will y the dollars with domestic currency, which tends to undo the initial moneyhtening operation. Therefore an active money or interest policy will eventuy require giving up either fixed exchange rates or capital mobility.

Under floating exchange rates, monetary policy can be stronger (Frenkel and zin 1987). However, the interest rate still satisfies equation 4, so that for a en (exogenously determined) expected rate of devaluation, the domestic inest rate is unaffected by monetary policy. Thus a policy of tight money will t raise the domestic interest rate, but will result in output loss. Equation 4 ows that the domestic interest rate rises, point by point, with the expected rate devaluation. Thus a high domestic interest rate implies a high expected rate of valuation, which is a likely signal that the stabilization program is not edible.

In Dornbusch's (1976) extension of the Mundell-Fleming model under flexible change rates, the expected rate of devaluation becomes an endogenous vari-ble. He shows that a permanent cut in the money supply will temporarily ppreciate the domestic currency with respect to its long-run level. Conseuently, the cut in the money supply induces the expectation of future devaluaons, which, by equation 4, tends to increase the nominal interest rate. Hower, the increase is a temporary phenomenon; the greater the degree of price exibility, the shorter the period of increase. Because price flexibility is likely to high in countries undergoing anti-inflation programs, the temporary appreation of the currency does not seem to be an important factor for justifying igh interest rates.

Lower-Than-Expected Inflation Rates

If inflationary expectations could be manipulated by a skillful politician, inflaon would immediately decline, and no further short-run monetary stringency

would be required. A smaller expected rate of inflation would shift the *IS* curve to the left, resulting in some output loss, but the nominal interest rate would fall. Success in reducing inflationary expectations results in low interest rates. Consequently, tight money and high interest rates are likely to be more necessary when the public is reluctant to lower its expectations of inflation.

Consider a situation of tight money and downward-inflexible inflationary expectations in the short run. As a result of the tight money policy, ex post inflation turns out to be less than expected. The ex post real interest rate, r, is the interest rate that a borrower will have to pay in real terms. It is defined as

$$r = i - \pi. \tag{5}$$

It may be that, because the inflation rate is lower than expected, the public is quick to adjust downward its expectations about inflation. The tight money policy, however, makes the ex post real interest rate larger than the ex ante one, which implies unexpected redistribution of wealth from borrowers to lenders. This unplanned redistribution may become critical if expectations about inflation are slow to adjust. If real interest rates remain high in subsequent months, for example, the redistribution may cause the productive sector to deteriorate. Investors may have undertaken prudent investment projects in the expectation that the real interest rate would be low (given that the expected rate of inflation is high). However, a fall in the inflation rate by more than expected for an extended period of time may drive all or most firms into bankruptcy.

Thus, if the management of expectations relies on stringent short-term monetary policy, generalized bankruptcy may result. A state of generalized bankruptcy implies that banks would have to terminate most of their loans, revealing the banks' weak financial position. Firms and banks may unwittingly be led into a scheme to hide these facts from the public and the regulators. As signs of trouble develop, the authorities may opt to loosen monetary policy. To solve bankruptcy problems, however, the ex post real interest rate must be lowered to about its ex ante levels, which would mean a return to high inflation. The longer the time that elapses between the initiation of the stabilization process and the return to high inflation, the higher will be the level of inflation required.

The high-inflation solution may be postponed by increasing the remuneration of bank reserves held against deposits to expand the money supply. This would be equivalent to increasing the nominal interest rate on money. The increase in the money supply would result in an increase in loanable funds channeled through the banking system and would help to keep inflation under control. This emergency policy cannot be maintained indefinitely, however.

III. Sustainability Issues

A policy of high interest rates may give rise to a state of generalized bankruptcy, which may impair the stabilization program in the medium run. In the short-run model, however, explicit attention was not paid to private or public

udget constraints. Thus we were not able to articulate the medium-term diffi-
ulties in a completely rigorous fashion.

The impact of the rate of interest on bank deposits has played an important
ɔle in recent Latin American experiences. Bobember and Makinen (1983) pre-
ent a fascinating account of the destabilizing effects of indexed deposits in
Iungary after World War II.

Increased Demand for Money

For the sake of simplicity, bank money is assumed to be the only form of
lomestic money held by the private sector (the private sector holds no domestic
ash). However, individuals hold foreign exchange, f. Hence, assuming purchas-
ng power parity and a unitary international price level, the private sector's real
inancial wealth, w, is given by

$$w = x + f \tag{6}$$

vhere x stands for real monetary balances; in the notation of section II, $x =$
xp$(m - p)$. The rate of return of foreign exchange in domestic currency is the
xpected rate of devaluation and that of domestic money is the interest rate on
ıoney. The demand for money is assumed to be a function, L, of its opportunity
ost, and thus

$$x = L(\epsilon^e - i_m), \qquad L' < 0. \tag{7}$$

ı general, the demand for money ought to be a function of financial wealth, but
ıis term is omitted for simplicity.

If the interest rate on bank deposits, i_m, increases and the exchange rate is
ɔnstant, according to equation 7 the demand for money will rise. The public
ill try to shift their financial portfolios out of foreign currency and into domes-
c deposits (equation 6). If the government does not intervene, the shift into
ɔmestic deposits will put downward pressure on the exchange rate, thus caus-
ıg the domestic currency to appreciate. Because of purchasing power parity,
ıis portfolio shift tends to lower the price level. Consequently, an increase in the
ıterest rate on money immediately helps to reduce the inflation rate.

If the monetary authority does not react to the higher demand for domestic
oney, the exchange rate and price level fall to accommodate the higher money
emand. From the fiscal point of view, the increase in the interest rate on money
ɔes not bring about any additional resources. Given the expected rate of deval-
ıtion, the burden of servicing deposits goes up. Therefore, except for the very
ıort-lived inflation relief, the increased interest rates may cause greater fiscal
fficulties.

The monetary authority may intervene so as not to allow the exchange rate to
ɔpreciate (or depreciate) following a rise in the interest rate on money. As a
sult the central bank will accumulate foreign exchange according to

$$\dot{f} = L(\epsilon^e - i_m) - x_{\bar{0}} \tag{8}$$

where $\dot{f}$ is the change in foreign exchange and $x_{\bar{0}}$ denotes the stock of rea monetary balances before the policy of higher interest rates is put into effect. I the accumulation of reserves has positive side effects, the policy reaps som short-run benefits. However, if reserves earn no real interest, the net fiscal effec is that the initial capital inflow given by equation 8 adds nothing to revenue.

If the real interest rate earned on reserves is r^*, then taking equation 8 int account, the net cost of servicing bank deposits, C_N, becomes

$$C_N = L(\epsilon^e - i_m)(i_m - r^*) + r^* x_{\bar{0}} \tag{9}$$

where we assume initially $i_m = 0$. That is, the net cost of servicing bank deposit equals the direct cost, $i_m x$, minus interest on reserves accumulated as a conse- quence of raising the interest rate on deposits from 0 to a positive number which, recalling equation 8, is given by $(x - x_{\bar{0}})r^*$. Hence a necessary conditio for the fiscal situation to improve (compared with setting $i_m = 0$) is that th interest rate on money be less than the interest rate on reserves ($i_m < r^*$). This i a very stringent condition, given that in actuality the rate of interest earned o reserves is less than 1 percent a month. The general implication, therefore, i that although positive nominal deposit rates may be desirable from a fiscal poin of view, their optimal level, when authorities are committed to zero devaluation is unlikely to exceed 1 percent a month.

Endogenously Determined Rate of Devaluation

Zero devaluation is, of course, a very strong assumption. In practice policy- makers commit themselves to such a policy for only limited periods. When th rate of devaluation is an endogenous variable, a realistic assumption is that th government runs a real deficit, denoted by g (assumed constant for simplicity) Money is issued to cover the real deficit and any cost of servicing bank deposits. Therefore,

$$\dot{M} = Eg + i_m M - r^* E(x - x_{\bar{0}}) \tag{10}$$

where a dot over a variable denotes its time derivative. M is the nominal stock o money, and E is the exchange rate (the foreign price of output is assumed to b unity). The first term on the right side of equation 10 is the nominal value of th government's nonbank deficit, the second term is the interest on deposits, an the third term is the nominal value of the yield on the initial accumulation o reserves at the central bank. Dividing both sides of equation 10 by the exchang rate,

$$\dot{x} = g - (r^* + \epsilon - i_m)x + r^* x_{\bar{0}} \tag{11}$$

where ϵ is the actual rate of devaluation.

It is assumed that the private sector immediately adjusts its expectations t actual values (that is, it operates under perfect foresight). Thus,

$$\epsilon^e = \epsilon. \tag{12}$$

;onsequently, by equations 7, 11, and 12, and holding real money balances onstant ($\dot{x} = 0$),

$$0 = g - (r^* + \epsilon - i_m)L(\epsilon - i_m) + r^* \bar{x}_0. \tag{13}$$

'hus, for a given interest rate on money, the steady-state equilibrium condition ı equation 13 determines the rates of devaluation that are consistent with the ıodel. In equation 13 the only thing that matters is the difference between the ctual rate of devaluation and the interest rate on money. Hence, a higher ıterest rate on money brings about an increase in all equilibrium rates of evaluation (and inflation) that is exactly equal to the initial rise in the interest ıte on money.

Thus, under perfect foresight, an increase in interest rates for deposits will :sult in higher steady-state inflation. If, in addition, the real nonbank govern-ıent deficit is not affected by the policy on interest rates, then the steady-state quilibrium rate of inflation increases point for point with the deposit interest ıte. (For similar implications in models in which bonds take the place of ıterest-bearing money, see Sargent and Wallace 1981 and Calvo 1985.) This :sult underscores the importance of looking beyond the short run in evaluating ıterest rate policies. Contrary to our discussion in the previous section, here the igher interest rate does not give rise to bankruptcies. The end result, however, not very different: inflation goes up.

IV. The Role of an Active Interest Rate Policy

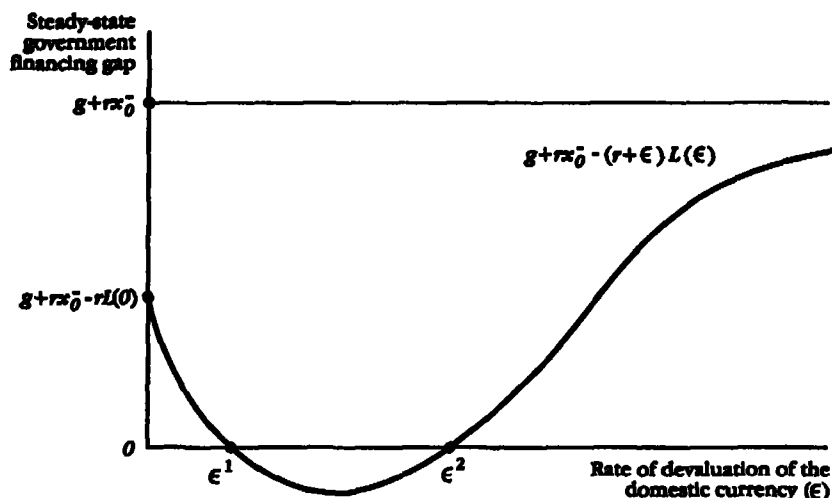
In this section we develop examples in which the economy displays equilib-um multiplicity and where there is room for an active interest rate policy.

Suppose the fiscal deficit is fully "monetized," as implied by equation 13. Even 'hen the rate of interest on money is zero, equilibrium may be consistent with ıore than one rate of devaluation. For instance, suppose the demand for money ı equation 7 satisfies the following:

$$L(\epsilon - i_m) = \exp[-\alpha(\epsilon - i_m)] \tag{14}$$

'hich corresponds to the Cagan model of the demand for money. Equation 13 ıay exhibit two solutions for the equilibrium rate of devaluation, as depicted in gure 2 at points ϵ^1 and ϵ^2.

The government could eliminate inflation indeterminacy by choosing the nallest equilibrium rate of devaluation (ϵ^1 in figure 2) and establishing a system f predetermined exchange rates. (See Bruno and Fischer 1985 on inflation ıdeterminacy and the "heterodox" program in Israel.) By controlling the rate of :valuation, the government can move the economy away from the high-interest ıd into the low-interest equilibrium. Although this is an example of an active ıterest rate policy, note that the optimal policy consists of pushing the system to *low*-interest equilibrium.

Figure 2. *Steady State Equilibriums*

Equilibrium Indeterminacy and Discretionary Policy

Equilibrium indeterminacy may be an even more serious problem if the government is likely to resort to discretionary policy. This point is illustrated by a simple example that also demonstrates the role that debt indexation may play (Calvo 1988 and 1989a). It is assumed that no interest is paid on money and that, contrary to the previous example, to finance the deficit the government resorts to issuing indexed and nonindexed bonds. Indexed bonds are a promise to pay the original value plus a return based on the real interest rate in units of output in the next period (the analysis is conducted in the context of a two-period model). Nonindexed bonds are nominal bonds that promise to pay the original value plus a return based on the nominal interest rate in units of domestic money in the next period.

Furthermore it is assumed that there is perfect capital mobility and that the international one-period real interest rate is zero. Therefore, in equilibrium the domestic real interest rate must equal zero. Likewise, if there is perfect foresight, in equilibrium the nominal domestic interest rate is equal to the rate of inflation. In equilibrium the implicit real interest rate on nonindexed bonds must also be equal to the international interest rate (which is zero by assumption).

To introduce discretionary policy in a simple manner, it is assumed that the rate of inflation cannot be determined until the next period. Moreover for simplicity we assume that in the next period the policymaker can choose the inflation rate to be either zero or some positive (possibly large) number ($\bar{\pi}$). Zero inflation has no social cost. If, however, a positive rate of inflation is implemented, the inflation cost is positive and, for simplicity, is assumed to be equivalent to enlarging the real government deficit.

Because this is a two-period world, total expenditure in the next period has to e completely financed by taxes. Taxes are distorting, and, again for simplicity, heir social cost is assumed to be proportional to tax revenue. It is assumed that he present government is stuck with a given amount of debt, which can, in rinciple, be split between indexed and nonindexed bonds (the sum of indexed nd nonindexed bonds, however, is given). Total next-period taxes (T_{t+1}) in real erms are given by

15a) $$T_{t+1} = g + B_i + B_n(1 + i), \qquad \text{if } \pi = 0$$

15b) $$T_{t+1} = g + B_i + B_n(1 + i)/(1 + \bar{\pi}) + \omega, \text{if } \pi > 0$$

vhere B_i is the stock of indexed bonds in terms of output, B_n is the stock of ionindexed bonds in terms of output, and ω is the social output cost associated vith high inflation. Equations 15a and 15b incorporate the equilibrium condi- ion that the domestic real interest rate must equal zero, since the return on ndexed bonds cannot be affected by the government in the next period.

Given that the rate of inflation can take only two admissible values, the only wo types of equilibriums are a low-inflation equilibrium and a high-inflation quilibrium. Under a low-inflation equilibrium, people at present expect infla- ion to be zero, and thus the interest rate is equal to zero. For this to be a perfect- oresight equilibrium, it is necessary for the government in the next period to ave the incentives to pursue a low-inflation policy, that is, to set the inflation ate to zero. Since the costs of each inflation policy are given by equations 15a nd 15b, it follows that setting the rate of inflation equal to zero would be the ptimal choice if equation 15b exceeds 15a when the interest rate is zero. Thus ie condition for low-inflation equilibrium is

16) $$\omega - B_n\bar{\pi}/(1 + \bar{\pi}) > 0.$$

In a high-inflation equilibrium, people expect the rate of inflation to be reater than zero ($\pi = \bar{\pi}$). The policymaker will have an incentive to validate xpectations of high inflation if equation 15a is larger than 15b. Thus the ondition for high-inflation equilibrium is

17) $$B_n\bar{\pi} - \omega > 0.$$

Conditions 16 and 17 are depicted in figure 3. If the stock of nominal bonds is qual to zero, then condition 17 can never hold, whereas condition 16 always oes. Hence the low-inflation equilibrium is the unique equilibrium. However, s shown in figure 3, there exists a critical value of nominal bonds, after which a igh-inflation equilibrium is possible. In the interval between B^1 and B^2 in figure , both low- and high-inflation equilibriums are possible, that is, the economy xhibits multiple equilibriums. After B^2 in figure 3 only high-inflation equilib- ium is possible, that is, only high inflation is a credible outcome.

Because in equilibrium the interest rate is equal to the rate of inflation, then, y equations 15a and 15b, the social cost of taxes and inflation is given by

Figure 2. *Steady State Equilibriums*

Equilibrium Indeterminacy and Discretionary Policy

Equilibrium indeterminacy may be an even more serious problem if the gov ernment is likely to resort to discretionary policy. This point is illustrated by a simple example that also demonstrates the role that debt indexation may pla (Calvo 1988 and 1989a). It is assumed that no interest is paid on money an that, contrary to the previous example, to finance the deficit the governmen resorts to issuing indexed and nonindexed bonds. Indexed bonds are a promis to pay the original value plus a return based on the real interest rate in units o output in the next period (the analysis is conducted in the context of a two period model). Nonindexed bonds are nominal bonds that promise to pay th original value plus a return based on the nominal interest rate in units of domes tic money in the next period.

Furthermore it is assumed that there is perfect capital mobility and that th international one-period real interest rate is zero. Therefore, in equilibrium th domestic real interest rate must equal zero. Likewise, if there is perfect foresight in equilibrium the nominal domestic interest rate is equal to the rate of inflation In equilibrium the implicit real interest rate on nonindexed bonds must also b equal to the international interest rate (which is zero by assumption).

To introduce discretionary policy in a simple manner, it is assumed that th rate of inflation cannot be determined until the next period. Moreover fo simplicity we assume that in the next period the policymaker can choose th inflation rate to be either zero or some positive (possibly large) number ($\bar{\pi}$). Zer inflation has no social cost. If, however, a positive rate of inflation is imple mented, the inflation cost is positive and, for simplicity, is assumed to be equiva lent to enlarging the real government deficit.

Because this is a two-period world, total expenditure in the next period has to be completely financed by taxes. Taxes are distorting, and, again for simplicity, their social cost is assumed to be proportional to tax revenue. It is assumed that the present government is stuck with a given amount of debt, which can, in principle, be split between indexed and nonindexed bonds (the sum of indexed and nonindexed bonds, however, is given). Total next-period taxes (T_{t+1}) in real terms are given by

$$T_{t+1} = g + B_i + B_n(1 + i), \qquad \text{if } \pi = 0 \tag{15a}$$

$$T_{t+1} = g + B_i + B_n(1 + i)/(1 + \bar{\pi}) + \omega, \text{ if } \pi > 0 \tag{15b}$$

where B_i is the stock of indexed bonds in terms of output, B_n is the stock of nonindexed bonds in terms of output, and ω is the social output cost associated with high inflation. Equations 15a and 15b incorporate the equilibrium condition that the domestic real interest rate must equal zero, since the return on indexed bonds cannot be affected by the government in the next period.

Given that the rate of inflation can take only two admissible values, the only two types of equilibriums are a low-inflation equilibrium and a high-inflation equilibrium. Under a low-inflation equilibrium, people at present expect inflation to be zero, and thus the interest rate is equal to zero. For this to be a perfect-foresight equilibrium, it is necessary for the government in the next period to have the incentives to pursue a low-inflation policy, that is, to set the inflation rate to zero. Since the costs of each inflation policy are given by equations 15a and 15b, it follows that setting the rate of inflation equal to zero would be the optimal choice if equation 15b exceeds 15a when the interest rate is zero. Thus the condition for low-inflation equilibrium is

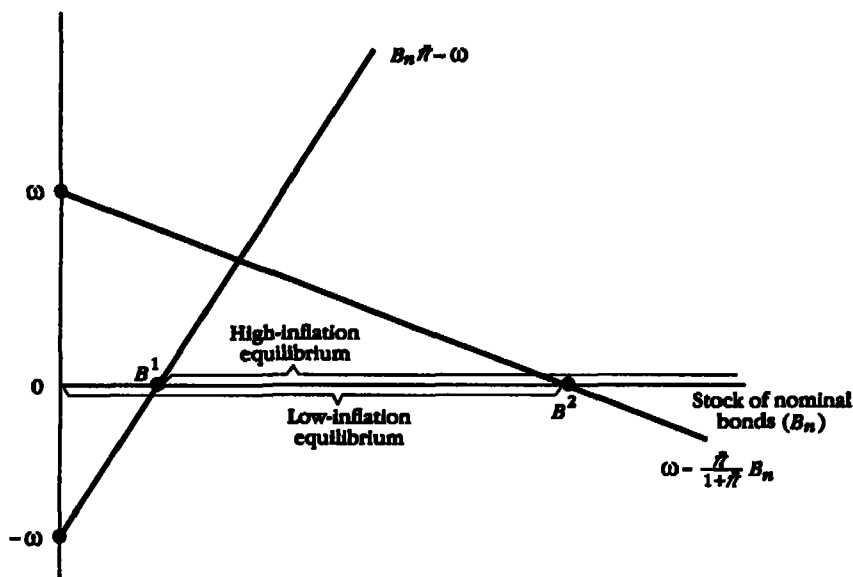
$$\omega - B_n\bar{\pi}/(1 + \bar{\pi}) > 0. \tag{16}$$

In a high-inflation equilibrium, people expect the rate of inflation to be greater than zero ($\pi = \bar{\pi}$). The policymaker will have an incentive to validate expectations of high inflation if equation 15a is larger than 15b. Thus the condition for high-inflation equilibrium is

$$B_n\bar{\pi} - \omega > 0. \tag{17}$$

Conditions 16 and 17 are depicted in figure 3. If the stock of nominal bonds is equal to zero, then condition 17 can never hold, whereas condition 16 always does. Hence the low-inflation equilibrium is the unique equilibrium. However, as shown in figure 3, there exists a critical value of nominal bonds, after which a high-inflation equilibrium is possible. In the interval between B^1 and B^2 in figure 3, both low- and high-inflation equilibriums are possible, that is, the economy exhibits multiple equilibriums. After B^2 in figure 3 only high-inflation equilibrium is possible, that is, only high inflation is a credible outcome.

Because in equilibrium the interest rate is equal to the rate of inflation, then, by equations 15a and 15b, the social cost of taxes and inflation is given by

Figure 3. *Equilibrium Determination and the Stock of Nominal Bonds*

Note: The stock of nominal bonds is nonindexed government debt.

$g + B_i + B_n$, if low-inflation equilibrium holds, and $g + B_i + B_n + \omega$, if high-inflation equilibrium holds. Hence low-inflation equilibrium is always socially preferable to high-inflation equilibrium. The intuition for this result is that, because of perfect foresight, the output value of the government's obligations will be the same irrespective of inflation. By assumption, the other items comprising government expenditure are independent of the rate of inflation. The only difference between a high- and low-inflation equilibrium is the social cost associated with high inflation, which proves that high-inflation equilibrium is inferior to low-inflation equilibrium.

Debt Indexation and Low-Inflation Equilibrium

From the above example, debt indexation can be helpful to ensure that the economy goes to the low-inflation and low-interest-rate equilibrium. The government can achieve a lower stock of nominal bonds by swapping indexed for nonindexed debt until the stock of nominal bonds becomes smaller than the critical level (B^1 in figure 3). (See Guidotti and Kumar 1991 on the use of indexed debt in the recent successful anti-inflationary programs in Chile and Israel.) The example's simplicity, however, hides the fact that inflation is not the only tool governments can employ to partially repudiate their debts. Domestic debt, for example, can be "taxed away" without resorting to higher inflation. Although indexation removes inflationary incentives, it may result in other

forms of debt repudiation that might be even more costly. In practice, therefore, the optimal policy will depend on specific country considerations, like the stability of private ownership legislation.

Multiple Equilibriums and Expectations about Inflation

It may be that, for considerations outside the model, debt indexation is not possible or desirable. If a low-inflation equilibrium prevails, there is nothing the policymaker could or should do. Much more interesting is the case in which the stock of nominal bonds lies in the range of multiple equilibriums. For if the public expects high inflation, interest rates will be high, and the government in the next period will have incentives to engineer high inflation. But if the public expects zero inflation, interest rates will be low, and low inflation will be the incentive-compatible outcome. In that context the policymaker could lead the economy to the best equilibrium by persuading the public that low inflation will be the final outcome. For that purpose it would be enough if the government refused to sell bonds with an (implicit or explicit) nominal interest larger than zero.

Even if the policymaker succeeds in moving the system toward low-inflation equilibrium, there is another potential complication. If the rate of inflation is zero and the stock of bonds is near B^2 in figure 3, then a slight increase in the stock of nominal bonds would imply that the system would be suddenly thrown into the high-inflation equilibrium. This indicates that the stabilization program may be very susceptible to the level of outstanding nonindexed government debt. More important, this kind of example shows that a lack of trouble spots, and even apparent success—for instance, a dramatic initial fall in the rate of inflation—are not necessarily reliable signs that the stabilization program is running on solid ground.

V. Recent Stabilization Programs in Eastern Europe

The policy of high interest rates has received renewed attention in recent stabilization programs in Eastern Europe and has been tried during the first few months of the 1990 stabilization program in Poland. One rationale for adopting high interest rates is that prereform inflation in some of these economies (notably Poland) was closely associated with the existence of substantially negative real interest rates. Therefore high interest rates were expected to send a strong signal about a change in the monetary regime.

Such a policy may be useful particularly when, as in Poland, it is carried out for a limited period, and enterprises start the program with excessive inventories (see Calvo and Coricelli, forthcoming). Indeed, a policy of high interest rates could dominate credit rationing—even if credit flows turn out to be the same as under high interest rates—because credit rationing has already been tried in the

past, and thus its adoption could fail to convey the impression that a dras policy change has taken place.

However, credit markets in Eastern Europe are notoriously segmented a incomplete. The productive sector strongly depends on central bank cred Even when interenterprise credit exists, it relies very much on the presence of t central bank as a lender of last resort and tends to dry up when the central ban withdraws its implicit support of problem borrowers. Therefore a policy of hig interest rates could have a direct negative effect on output. Aside from weaken ing political support for the program, which may be caused by excess capacity and unemployment, output contraction has a direct negative impact on fiscal revenue, weakening policy sustainability and, thus, the program's credibility (Calvo and Coricelli 1991).

VI. Conclusion

High interest rates may jeopardize the success of an inflation stabilization program, and, if the program does not enjoy full credibility, the contraction that high rates are supposed to generate may not materialize (enough) in the short run. If high interest rates fail to achieve short-run adjustment, the adjustment may come later but at a much greater expense, because it may entail a state of generalized bankruptcy or a demise of the stabilization program.

The policymaker's dilemma is not an easy one. On the one hand, high interest rates send a clear signal that a new regime has been put in place. Banking subsidies are reduced or eliminated, and savers have a bigger incentive to make their funds available for productive investment. Furthermore, in the short run high interest rates may decrease aggregate demand, helping the stabilization effort. On the other hand, as shown here, high interest rates may cause the program to fail.

Whenever credibility issues are stake, the institutional and historical characteristics of the country in question play a central role (see Calvo 1989b). It is relatively easy to point to policies—such as a policy of high interest rates—that have a good chance of failing. A more positive suggestion that comes out of the analysis is that interest rates should be set in such a way that they do not contradict the sustainability of the program. A less ambitious stabilization program should be implemented. A program should not plan on real interest rates exceeding 15–20 percent a year for an extended period of time, unless there are very good reasons to argue that those rates are consistent with the country's available investment projects.

This quantitative constraint on real interest rates is likely to be hard to satisfy in highly indebted countries because of debt overhang and related problems. In such countries financial liberalization will be slower. The binding constraint is not an undersized banking system, but a lack of investment possibilities given the uncertainties created by a large outstanding debt. The country has a debt problem, and it is therefore unlikely that more debt would be the solution.

References

ord "processed" describes informally reproduced works that may not be com-
available through library systems.

nber, William A., and Gail E. Makinen. 1983. "The Hungarian Hyperinflation Stabilization of 1945–46." *Journal of Political Economy* 91 (October): 801–24.

o, Michael, and Stanley Fischer. 1985. "Expectations and the High Inflation Trap." IT, Cambridge, Mass. Processed.

vo, Guillermo A. 1985. "Macroeconomic Implications of the Government Budget." *urnal of Monetary Economics* 15 (January): 95–112.

——. 1988. "Servicing the Public Debt: The Role of Expectations." *American Economic Review* 78 (September): 647–61.

——. 1989a. "Controlling Inflation: The Problem of Non-Indexed Debt." In Sebastian Edwards and Felipe Larrain, eds., *Debt Adjustment and Recovery: Latin America's Prospect for Growth and Development*. New York: Basil Blackwell.

——. 1989b. "Incredible Reforms." In Carlos Diaz-Alejandro and Guillermo Calvo, eds., *Debt, Stabilization, and Development*. New York: Basil Blackwell.

alvo, Guillermo A., and Fabrizio Coricelli. 1991. "Stagflationary Effects of Stabilization Programs in Reforming Socialist Countries: Enterprise-Side and Household-Side Factors." *The World Bank Economic Review*, this issue.

——. Forthcoming. "Stabilizing a Previously Centrally Planned Economy: Poland 1990." *Economic Policy*.

rnbusch, Rudiger. 1976. "Expectations and Exchange Rate Dynamics," *Journal of Political Economy* 84 (6, December): 1161–76.

nkel, Jacob A., and Assaf Razin. 1987. "The Mundell-Flemming Model a Quarter Century Later." IMF *Staff Papers* (December): 567–620.

idotti, Pablo, and Manmohan Kumar. 1991. "Domestic Public Debt of Externally ndebted Countries." IMF Occasional Paper 80. Washington, D.C.: International Montary Fund.

cks, John R. 1937. "Mr. Keynes and the Classics: A Suggested Interpretation." *Econometrica* 5: 147–59. Reprinted in W. Fellner and B. F. Haley, eds., *Readings in the Theory of Income Distribution*. London: George Allen and Unwin, 1961.

driguez, Carlos A. 1988. "Comment." In Michael Bruno, Guido Di Tella, Rudiger Dornbusch, and Stanley Fischer, eds., *Inflation Stabilization*. Cambridge, Mass.: MIT Press.

gent, Thomas J., and Neil Wallace. 1981. "Some Unpleasant Monetarist Arithmetic." *Quarterly Review of the Federal Reserve Bank of Minneapolis* (Fall): 1–17.

: WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 71-90

Stagflationary Effects of Stabilization Programs in Reforming Socialist Countries: Enterprise-Side and Household-Side Factors

Guillermo A. Calvo and Fabrizio Coricelli

Dismantling subsidies could give rise to serious macroeconomic difficulties in the short run. This article explores a view based on the enterprise sector as a central source and main channel of the stagflation phenomenon, using as an example the stagflation that followed the 1990 stabilization program in Poland. The stagflation phenomenon is linked to features of the financial market that are somewhat peculiar to reforming socialist economies: the weak credit links between households and enterprises, and the existence of large interenterprise debt. The policy implications of the enterprise-side view include more explicit consideration of initial conditions in the credit market, implementation of privatization schemes, and the development of a domestic banking system.

'eral socialist countries are in the process of transforming their economies into ystem that relies much more heavily on market mechanisms. A pervasive ıracteristic of the associated transformation programs is the elimination of :e distortions through dismantling a complicated network of subsidies. This ıspired by the valid microeconomic principle that an undistorted price system ıtributes to efficient resource allocation in the long run. In the short run, ıever, the elimination of subsidies could give rise to serious and protracted croeconomic difficulties. For example, eliminating subsidies could bring ›ut a sharp increase in the price of some key factors of production, such as ıl and gasoline. Thus, if firms have limited access to credit or if interest rates high, the increase in production costs could lead to a credit crunch. This icle pays special attention to the recent Polish stabilization plan. The Polish n exhibits a sizable initial increase in production costs, and its relative long- ty helps to distinguish a little better between blips and trends. (A more com-

uillermo A. Calvo is in the Research Department at the International Monetary Fund, and Fabrizio icelli is in the Country Economics Department at the World Bank. The authors would like to ıowledge the useful comments by Eduardo Borensztein, Luis de la Calle, Roman Frydman, Stanislaw nulka, Peter Hole, Timothy Lane, Massimo Russo, Andrés Solimano, Ulrich Thumm, and three ıymous referees and to thank Barbara Ossowicka for her efficient research assistance.

plete evaluation of the Polish program is taken up in Calvo and Coricelli [1991].)

An initial stagflationary period accompanied the Polish stabilization program. This article explores a view based on the enterprise sector as a central source and main channel of the stagflation phenomenon. It is suggested that economic transformation programs of the Polish type may contain a significant supply-side shock stemming from the combined effects of a steep rise in input prices (both domestic and foreign) and interest rates, together with an expansion of credit that fails to keep pace with input prices. The "material" intensity of production, namely the use of inputs and energy per unit of output, is extremely high in socialist economies (Gomulka and Rostowski 1988). This implies that an increase in input prices has a particularly large impact on the costs of enterprises. As a result, firms in Poland curtailed their demand for other firms' products and, eventually, their demand for labor. Thus the view discussed here does not preclude demand by enterprises to play an important contractionary role.

In this context the withdrawal of the central bank from its role as "lender of last resort" (seen by the authorities as a necessary step for the switching from a "soft" to a "hard" budget constraint) may have magnified the credit squeeze. Also, given the initial credit segmentation, which delinks the current profitability and liquidity conditions of enterprises from their efficiency, such an attempt at introducing a hard budget constraint may have adversely affected firms in a manner largely unrelated to their efficiency. In this regard the presence of a sizable stock of interfirm debt in the period preceding the stabilization program—probably with good firms in the position of creditors—may have played an important role in further spreading across firms the recessionary effects of the liquidity crunch.

There are similarities between the enterprise-side view and the neo-structuralist approach applied to nonsocialist developing countries (see, among others, Cavallo 1977, Taylor 1980, and van Wijnbergen 1982 and 1983). As in the neo-structuralist approach, the enterprise-side view stresses the possible contractionary effects of tight credit in the presence of serious imperfections in the financial market. However, the view exposed here links the stagflation phenomenon to features of the financial market that are somewhat peculiar to reforming socialist economies: the weak credit links between households and enterprises, and the existence of large interenterprise debt. The latter, in turn, is linked to chronic "soft budget constraints" and the above-mentioned financial separation of households and firms during the prereform period. In addition, as shown in the appendix, we emphasize the effects of stock liquidity and make the central points in terms of a model with perfect price flexibility. In contrast, the neo-structuralist school relies on flow effects and Keynesian price-wage rigidities.

Section I identifies the main relevant stylized facts of the Polish experience. Section II sketches some of the features of the financial market in reforming socialist economies. Section III lays out the enterprise-sector view of these facts,

ing on the simple model contained in the appendix. Section IV contains some icy *implications.*

I. The Polish Stabilization Program

On January 1, 1990, the Polish government launched a far-reaching program f economic transformation and stabilization. On impact, the inflation rate ccelerated sharply and industrial output fell (figures 1 and 2). The initial effects f the program can be characterized as stagflationary. Although official figures ay overstate the fall in output (Osband 1991), both the increase in the inflation te and the drop in production significantly overshot official predictions. This agflationary outcome is of particular interest because the Polish program can e classified as an exchange-rate-based, heterodox program. As shown by iguel and Liviatan (1990), this type of program is usually accompanied by itial expansion and very seldom results in a deep recession, such as occurred in oland. This comparison may not be fully warranted because the Polish pro-ram entailed a major structural transformation. However, although the struc-ural transformation may call for a fall in output, this would occur only if new rices are lower than new average costs. Otherwise output will not fall, and the

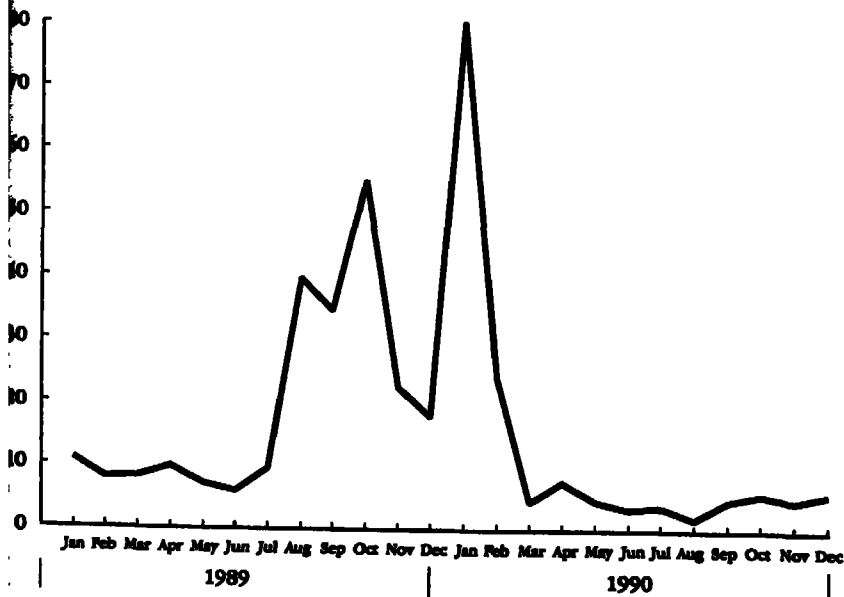
igure 1. *Inflation in Poland, 1989–90*
(Monthly percentage change in consumer prices)

Source: Central Statistical Office (1990).

Table 1. *Real Wages in Poland, 1989 and 1990*

Month and year	*Index of real wages (January 1989 = 100[a])*	*Wages in U.S. dollars[b]*
1989		
January	100	20
March	146	38
June	125	26
September	132	27
December	138	82
1990		
January	79	67
February	73	76
March	99	105
April	84	97
May	77	91
June	76	93
July	81	105
August	83	110
September	86	119
October	92	135
November	98	146

Note: Data are for average monthly wages and include profit bonuses.

a. Deflated by CPI inflation.

b. Average wage in zlotys converted into U.S. dollars at the free market exchange rate.

Source: Authors' calculations based on data from Central Statistical Office (1990) and National Bank of Poland (1990).

of domestic money, and in February 1990 it amounted to more than 60 percent of the M3 money supply. Second, although the fall in real wages is substantial (about 30 percent), the index number that measures this change does not take into account the widespread shortages in Poland in 1989. Therefore the effective real wage (or the real-real wage, as some Polish economists like to call it) in 1989 could have been much lower than the one implied by available statistics. For instance, if measured in terms of dollars—using the black market exchange rate as a (undoubtedly rough) proxy for shortages in the goods market—wages increased substantially during 1990 (see table 1). Third, other things being equal, a fall in workers' incomes should be associated with a rise in firms' profits by the same amount. In socialist economies workers are shareholders, either directly as recipients of profit bonuses or indirectly as beneficiaries of state surpluses. Thus one should expect a less pronounced reaction to a wage cut than if wage income were the only source of noninterest income of the household sector.

The household-demand view overlooks the importance of the supply shock provoked by the January increase in administrative input prices. As shown in table 3, the magnitude of the increase in input prices was extraordinary. Furthermore, at the beginning of 1990 credit to firms (plus domestic liquid assets held by firms) per unit of cost fell by about 60 percent compared with its average during 1989. Another indication of the liquidity squeeze of state enterprises is that wages at those firms stood below the program's ceilings for the first quarter of 1990.

Table 2. *Deposits of the Household and Enterprise Sectors in Poland, December 1989–August 1990*

Month and year	Household sector: Total deposits (trillion zl.)	Household sector: Foreign currency deposits[a] Trillion zl.	Household sector: Foreign currency deposits[a] Billion US$	Household sector: Foreign currency as a percentage of total deposits	Enterprise sector: Total deposits (trillion zl.)	Enterprise sector: Foreign currency deposits[a] Trillion zl.	Enterprise sector: Foreign currency deposits[a] Billion US$	Enterprise sector: Foreign currency as a percentage of total deposits
1989								
December	8.6	36.5	4.9	80.1	7.8	15.5	2.3	66.5
1990								
January	11.0	45.2	4.8	80.4	15.2	14.1	1.5	48.1
February	13.6	46.3	4.9	77.3	17.0	11.2	1.2	39.7
March	16.5	47.8	5.0	74.3	19.7	9.8	1.0	33.2
April	18.3	49.3	5.2	72.9	22.7	8.6	0.9	27.5
May	20.3	50.4	5.3	71.3	27.6	7.6	0.8	21.6
June	22.7	50.1	5.3	68.8	31.4	6.7	0.7	17.6
July	25.2	51.0	5.4	66.9	37.9	6.2	0.6	14.1
August	27.6	52.4	5.5	65.5	44.3	5.9	0.6	11.8

a. Foreign currency deposits are valued at the free market exchange rate.
Source: National Bank of Poland (1990).

Table 3. *Changes in Administrative Input Prices in Poland, 1986–90*

	1986[a]	1987[a]	1988[a]	October 1989[b]	January 1990[b]
Natural gas to industrial users				100.0	—
Crude petroleum and products[c]	31.2	39.0	57.0	80.0	—
Hard coal to industrial users	22.1	36.7	60.4	100.0	400.0
Electricity to industrial users	18.2	36.6	50.0	142.0	300.0
Overall producer price	17.8	26.6	59.8	43.9	109.8

— Not available.
a. Percentage change over previous year.
b. Percentage change over previous month.
c. Heavy oil for 1986–88.
Source: Central Statistical Office (1990).

There are two key financial aspects of the January stabilization program. One relates to the initial conditions characterizing the enterprise sector at the outset of the program. In this regard, we stress the importance of the interfirm credit market, which had flourished in previous years to reach a stock value of about 45 percent of GDP at the end of 1989. The other key aspect relates to a fundamental change in the role of the central bank. In particular, the January program sent a clear signal that the central bank was withdrawing from its role as lender of last resort. This is a fundamental institutional change, which may have strong consequences on the functioning of the interfirm credit market. Before this institutional change took place, interfirm credit received the implicit full backing of the central bank. If a firm ran into financial difficulties and could not service its debt to other firms, the central bank provided the necessary funds. In the new regime, by contrast, the central bank is expected to be much less accommodating. As a result lenders now have much stronger incentives to learn about the financial health of potential borrowers.

Information about creditworthiness is likely to have been scant at the beginning of the plan for two reasons. First, when there was backing from the central bank, information about borrowers' creditworthiness was much less valuable to the lender. Second, the new rules of the game were not well known, and, in particular, it was too early to know which firms would survive and which ones would have to be phased out. As a result private credit markets may have been especially slow to react to the credit squeeze, which may have magnified the negative impact of the credit crunch on output and employment and contributed significantly to the persistence of these adverse effects.

II. Financial Markets in Planned Socialist Economies: An Overview

The financial system of classical centrally planned economies may be simply characterized as being composed of highly separated monetary circuits: household money and enterprise money. Enterprise money is passive in the sense that monetary flows in the enterprise sector adjust automatically to planned real flows. The central planner sets targets on real flows and, accordingly, extends credit to enterprises for their needs, taking into account administered wages and

prices. In practice the banking system is monolithic, with the central bank directly administering all the monetary transactions of enterprises. Money tends to be active in the household sector in the sense that budget constraints precede household actions (consisting mostly of consumption decisions) in the manner assumed by standard microeconomic theory.

In a reformed planned economy, enterprises gain some independence about their decisions. They have some degree of freedom in determining wages, prices, input demand, and investments. Enterprise money can still be passive if the central bank accommodates all the actions of the enterprises, but credit accommodation is not a necessary feature of the system; it is, in principle, a policy decision of the central bank. Because bankruptcy is ruled out, however, attempts to curtail central-bank credit are quickly followed by a compensatory expansion of interfirm credit. The compensatory adjustment is relatively quick to materialize because interfirm credit bears little risk. The central bank is de facto forced to operate as lender of last resort (Kornai 1980) and is unable to control the supply of liquidity. As a result, stabilization programs tend to fail. Moreover repeated attempts to lower inflation may solidify the interfirm credit market. Thus stabilization programs launched under these no-bankruptcy conditions may induce a higher inflation plateau in the future and reduce even further the ability to control the supply of liquidity.

Reformed planned economies still tend to be characterized by strongly segmented credit markets and very weakly integrated monetary circuits of households and enterprise sectors. However, unlike centrally planned economies, the development of a thick interfirm credit market allows state enterprises in reformed planned economies to escape official credit ceilings. The ability of the firms to circumvent credit ceilings results in a loss of monetary control by the monetary authority.

Interfirm credit plays an important role in the reformed planned economies of Hungary (since the 1970s), Yugoslavia, Poland, and China. In Yugoslavia the stock of interfirm credit was estimated at 43 percent of gross social product in 1987; in Poland at around 48 percent of GDP at the end of 1989. In China the phenomenon is particularly relevant, and in the package of policy measures implemented by the Chinese government in 1989, the solution to this problem was deemed as one of the main targets. It is estimated that, in 1990, arrears in payments of interfirm credit amounted to about 20 percent of the working capital of all state enterprises, and arrears in payments of working capital credits from the banking system were of analogous size.

III. The Enterprise-Side View of Stagflation

Despite its oversimplified structure, the enterprise-side explanation appears able to account for the type of stagflation that occurred in Poland after the stabilization program was implemented. The recent stagflation may be partly due to the substantial increase in input prices, which may have been magnified by the presence of segmented credit markets. In particular there seems to be a

severe segmentation between the households and firms. A model of the enterprise-side view of stagflation is presented in detail in the appendix.

In a more realistic model some of the fall in output could actually reflect an efficient reallocation of resources. For example, before the stabilization program, firms may have overaccumulated inventories and taken advantage of the highly negative real interest rates. Thus the fall in output of firms producing inventory may actually lead the system to a socially better equilibrium. Consequently, the welfare implications of a credit crunch require further scrutiny (Calvo and Coricelli 1991).

Model Assumptions

Consider an economy of identical (Ramsey-type) individuals whose utility function depends on present and future consumption streams. Households are assumed to have perfect access to international capital markets, where they can freely borrow and lend at the international rate of interest. This is an extreme assumption; however, it corresponds to a situation not too different from the one prevailing in Poland. Polish households carry in their portfolios a sizable stock of foreign currency deposits. Therefore at any time they are free to lower those balances to buy, for example, durable goods. This is equivalent to borrowing in international markets because—like borrowing—it results in a fall of the net financial position of Polish households in relation to the rest of the world. Furthermore, if those funds are held in accounts yielding the international risk-free interest rate (which is approximately the case in Poland for dollar-denominated bank deposits), then the cost of those "newly borrowed" funds corresponds to the interest rate assumed in the model.

Firms are assumed to produce a tradable consumption good by means of an imported intermediate input (for example, coal or oil) and a fixed amount of labor. Firms are price takers but wage makers; firms are controlled by workers' councils, which are assumed to maximize the present discounted value of wages (evaluated at the international interest rate). This is a natural objective of firms in the present context, given the previous assumption that workers have perfect access to international capital markets. The key assumption of the model, however, is that firms face a constraint of needing liquidity in advance to finance intermediate inputs. This means that in order to finance a given flow of expenditure, firms must secure a stock of liquidity that must bear a given "technical" relationship to the flow of expenditure. Liquidity at a given firm is defined as the money it holds plus credit lines from the central bank and suppliers available to the firm. In this fashion, firms face a liquidity constraint that has to hold at each point in time.

The Case When Firms Are Unable to Borrow

In the simple case in which firms are unable to borrow from either the central bank or the private sector (that is, other firms and households), the liquidity of the firms is entirely given by the stock of money they hold. At each point in time the stock cannot be changed unless the firms sell inventories (which is ruled out

in this simple version of the model). Thus, if firms start from a position in which the liquidity constraint is binding, then a one-step rise in the price of the input may provoke a liquidity crunch and an immediate output loss. The output loss occurs irrespective of what happens with output prices. This observation shows that a liquidity crunch can develop even when output prices increase in the same (or larger) proportion as input prices.

Firms will try to offset the liquidity crunch by drastically lowering real wages until a level of liquidity consistent with full-capacity utilization is restored. Consequently, output will fall on impact and then steadily climb back to its higher, long-run level. Profits—which are equal to the accumulation of real monetary balances held by firms—will display a dramatic increase after the credit crunch. Thus, in the midst of this recession driven by the credit crunch, profits of firms will be spectacular. A fall in real wages is consistent with the Polish experience in early 1990 (see table 1), and it helps to explain why wage ceilings were not binding at the beginning of the program (Frydman and Wellisz 1991).

The high profitability of enterprises could easily lead to the wrong conclusion *that there is no credit problem. However, high profitability actually reflects the* attempt by firms to relieve the credit crunch. Consequently, it would be somewhat unwarranted to interpret the unexpectedly high profitability of the socialized sector in Poland during January–April 1990 as a clear indication that credit does not lie at the heart of the recession.

The fall in output implies a decline in permanent household income, which is, however, smaller than the fall in transitory household income. Income is expected to return to its full-employment level. Thus, since consumption is determined by permanent income, this version of the model predicts an initial trade balance deficit, as the decline in production is, on impact, larger than the decline in consumption. However, in Poland following the stabilization program there was a marked improvement in the trade balance. In 1990 it is estimated that the trade balance showed a surplus of US$2.7 billion compared with US$0.2 billion in 1989.

The model can easily be modified to account for a temporary improvement in the trade balance. For example, enterprises may hold sizable stocks of inventories at the start of the program—partly, perhaps, in anticipation of the forthcoming liquidity crunch. Thus, in response to the liquidity crunch, firms are likely to react by using up their own inventories and lowering their demand for new intermediate goods. Therefore the liquidity crunch could lead to a reduction of total absorption (consumption by households and absorption by firms) that exceeds the fall in output, yielding a surplus in the balance of trade. In this case the fall in output is reduced because inventories act as a buffer stock.

The Case with Heterogeneous Firms

The model can be further enriched by assuming that firms are heterogeneous and that some firms produce intermediate goods used as factors of production by other firms in the economy. In a reformed planned economy, firms have

incentives to lend to one another in the expectation that the monetary author will bail out firms that develop financial problems. Therefore, if the reforme planned economy launches a stabilization plan similar to the one in Poland in 1990, the economy is likely to start the program with a positive—and possibly large—stock of interfirm debt.

In such a framework, consider once again the effects of a rise in the price of primary imported inputs. In addition to the adjustments implied by previous versions of the model, it is now possible for a debtor firm to alleviate the credit squeeze by falling into arrears with other firms. The cost of falling into arrears depends on the effectiveness and credibility of bankruptcy regulations. If such regulations are believed to be less than fully effective, then firms are likely to find it attractive to fall into arrears with other firms. This, in and of itself, will increase the riskiness of the interfirm credit market and dry up sources of funds for solvent and efficient firms. Furthermore in the short run the buildup of arrears plays against firms that are net creditors at the start of the stabilization program. This is worrisome, because there are no a priori reasons to expect the more efficient firms to emerge unscathed from these financial squabbles.

The Official Banking Sector

The assumption of no official bank credit is, of course, unrealistic. However, the model could easily be extended to incorporate an official banking sector that takes deposits from households and lends to enterprises. This sector played a prominent role in Poland and was the main source of the easy money that prevailed until recently. By setting credit ceilings for the entire banking system, the Polish stabilization program limited the amount of credit that banks could offer to enterprises. In addition the refinancing rate (that is, the interest rate charged by the central bank to the other banks in the system) was raised considerably during the first few months of the program (table 4) and became highly positive in terms of foreign exchange. Given the imperfections of the banking sector—and, in particular, the above-mentioned segmentation between house-

Table 4. *Interest Rates in Poland, December 1989–December 1990*

Year	*Month*	*Refinancing rate*	*Average prime rate on loans*
1989	December	11.7	14.1
1990	January	36.0	46.5
1990	February	20.0	22.0
1990	March	10.0	10.5
1990	April	8.0	8.5
1990	May	5.5	6.3
1990	June	4.0	4.5
1990	July	2.8	2.9
1990	August	2.8	2.9
1990	September	2.8	2.9
1990	October	3.6	3.4
1990	November	4.6	4.2
1990	December	4.6	5.1

Source: National Bank of Poland (1990).

holds and enterprises—the central bank's increase of its interest rate induced banks to offer lending interest rates well above the refinancing rate. This interest rate policy seems to account for the fact that the credit ceiling was not binding in early 1990.

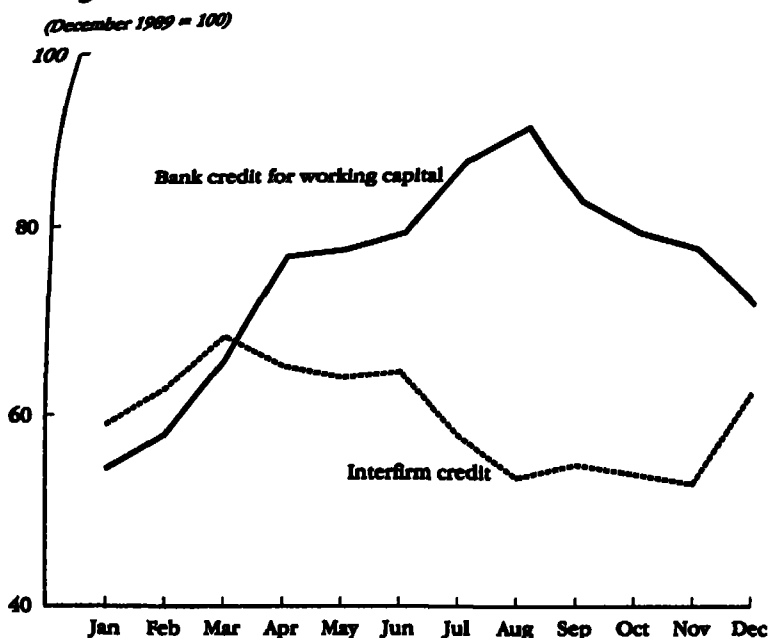
The excess supply of credit may appear to contradict the argument of a credit runch. However, the fact that the banks' interest rates were so closely geared to he refinancing rate of the central bank demonstrates that domestic credit mar-cets are highly imperfect. Therefore it is perhaps a good approximation to nvision the central bank as a credit monopolist able to set the price of credit. Thus, by raising the refinancing rate, the central bank would elicit a lower lemand for credit and, furthermore, could even ensure that any credit ceiling vould not be binding (assuming that firms would not expect to be fully bailed out). Under these circumstances the central bank could generate a credit crunch ither by drastically lowering credit ceilings or, alternatively, by sharply increas-ng the refinancing rate. Consequently, the view that credit ceilings were too ight is not essentially different from the view that interest rates were set too igh. The latter view is equivalent to saying that the virtual credit ceiling (as pposed to the actual credit ceiling) was too tight.

Credit tightness—measured in terms of bank credit—seems to have eased in he second half of 1990. However, output has responded much less to credit xpansion than to the credit contraction of the first half of the year. This is a uzzling development for the simple model outlined above and in the appendix. One explanation may be that the objective of firms is to maximize current wages ather than permanent wages, as postulated in the model. Thus, the easing of redit conditions would lead firms to offer higher wages instead of using the roceeds to relieve the liquidity squeeze. This seems consistent with the sharp ncrease in real wages in the second half of 1990 (Calvo and Coricelli 1991).

Another explanation is that a credit crunch signifies a fundamental shakeup of nterfirm relationships, leading firms to fall into arrears with one another and to reak all sorts of explicit and implicit contracts. In such an environment, there-ore, a later expansion of credit may not be able to mend the damage caused by he initial credit crunch, thus leading to a slower recovery. Furthermore some of he arrears with other firms will be repaid. Therefore easing up bank credit vould lead to a repayment of debts to both workers and suppliers, rather than n expansion of output. A measure of credit conditions that encompasses inter-nterprise credit as well would be more significant in this context, as bank credit nd interenterprise credit would be substitutes for each other. Some evidence of his ability to substitute for each other is given in figure 3, which shows that the xpansion of bank credit in the third quarter was matched by a contraction of nterenterprise credit, leaving total credit for enterprises roughly unchanged.

IV. Some Implications of the Enterprise-Side View

A credit crunch is likely to happen if the stabilization program starts with a arge devaluation of the currency and removal of subsidies on primary (tradable)

Figure 3. Interfirm Credit and Bank Credit in Poland, 1990

Note: End-period stocks deflated by the consumer price index.
Source: National Bank of Poland (1990).

commodities. If excessive credit tightness occurs (that is, if the fall in output exceeds what can be accounted for by considerations of efficiency) there are, in principle, three types of solutions: appreciate the currency, phase out the input subsidies more gradually, or increase central bank credit to the state enterprise sector. Appreciating the currency is the most attractive solution, because it would also lower the cost of primary inputs—thus directly relieving the credit crunch—and could most easily be presented to the public as a technical adjustment of the stabilization program. The other two solutions may be taken as signals that the government is weak and ready to abandon the stabilization plan.

If central bank credit to the state enterprise sector is increased, the liquidity of firms should undergo a stock jump. If the increase occurs only in the flow of credit to the enterprise sector, there is likely to be a suspicion that the government has gone back to its old inflationary practices. In contrast, the stock increase need happen only once and could possibly be disguised under a currency reform.

A stabilization program should consider initial conditions in the credit market. In this regard the initial stock and distribution of interfirm debt is of

paramount importance. Special attention should be given to avoid transmitting the financial difficulties of inefficient enterprises to efficient enterprises.

Demand-side solutions may be counterproductive. Forcing firms to pay higher wages exacerbates the credit crunch directly, and increasing government demand for goods and services leaves output unchanged and is likely to worsen the balance of payments. There were some indications of this phenomenon in Poland at the end of 1990, when the easing up of fiscal policy was accompanied by both a worsening balance of payments and an accelerating inflation.

Proposals for privatization are generally based on considerations of microeconomic efficiency. The enterprise-side approach suggests that, when foreign companies are involved, privatization can play a fundamental role even during stabilization programs. Foreign companies are likely to have better access to international capital markets, which would help to relieve tight domestic credit conditions. Clearly, besides the implementation of privatization schemes, the development of a domestic banking system has enormous relevance. This will indeed facilitate the transfer of funds among firms and between households and firms, especially when stabilization policies are credible (Calvo 1989).

Appendix. A Model of the Enterprise-Side View of Stagflation

The central formal ingredients of the model discussed in section III are sketched here. For simplicity, we will focus on the case in which all goods are tradable and there are no stocks of inventories. (Nontradable goods and other extensions are discussed in Calvo and Coricelli 1991.)

Households

Households are assumed to be homogeneous and to have a utility function:

$$\text{(A-1)} \qquad \int_0^\infty u(c_t)e^{-rt}dt, \qquad r > 0$$

where u is a strictly concave and twice-differentiable, instantaneous, utility function; c denotes consumption of the representative household; and r is its constant subjective rate of discount. The household's budget constraint is given by the following standard expression (Calvo 1986):

$$\text{(A-2)} \qquad \int_0^\infty c_t e^{-rt}dt = a_0 + \int_0^\infty (w_t + g_t - i_t m_t)e^{-rt}dt$$

where a, w, g, and m denote real (that is, in terms of consumption) financial wealth, labor income, government lump-sum transfers, and holdings of monetary balances corresponding to the representative household, respectively; moreover, i and r denote, respectively, the domestic nominal interest rate and the real (constant) international interest rate. It is assumed that the international real

rate of interest, r, is equal to the subjective rate of discount in expression A-1 Financial wealth, a, satisfies

$$a_t = m_t + b_t \tag{A-3}$$

where b is the stock of international bonds held by the representative household (for example, dollar-denominated deposits). The consumer is subject to a cash-in-advance constraint that takes the following form:

$$m_t \geq \alpha c_t, \qquad \alpha > 0. \tag{A-4}$$

The representative household is assumed to maximize its utility subject to its budget constraint; the liquidity constraint given by expression A-4 and its initial financial wealth, a, with respect to the path of consumption; and real monetary balances. If the government credibly announces that the exchange rate will be kept constant forever, then, by the assumption of perfect capital mobility, the domestic nominal interest rate equals the international rate. Therefore the solution to the above optimization problem calls for setting consumption at a constant level, and the liquidity constraint is binding.

Firms

Firms do not have access to the international capital market. They produce tradable consumption goods by means of a tradable primary input, n, and labor. For simplicity, labor employment at each firm is fixed and equal to one. Each firm is subject to a cash-in-advance constraint of the form:

$$\gamma p_t n_t \leq z_t, \qquad \gamma > 0 \tag{A-5}$$

where p denotes the price of the intermediate factor of production in terms of consumption, and z stands for real monetary balances held by the firm. Variable p is determined by the exogenously given international relative price of the primary good in terms of consumption ("small country" assumption), adjusted by local tariffs and taxes. Therefore output net of direct intermediate-input cost is given by

$$\Psi(n) - pn \equiv \phi(n; p) \tag{A-6}$$

where $\Psi(n)$ denotes the production function, which satisfies standard regularity conditions, and ϕ can be interpreted as the "net" production function. Net output is denoted by y; hence,

$$y_t = \phi(n_t, p_t). \tag{A-7}$$

We assume, without loss of generality, that the liquidity constraint is binding (that is, $\gamma pn = z$). In addition, we assume that the firm can only accumulate money balances. Therefore, assuming zero international inflation, the change in real monetary balances held by the firm is

$$\dot{z}_t = \phi(z_t/p; p) - \omega_t \tag{A-8}$$

where ω is labor income (or the wage) offered by the representative firm. Moreover, to simplify, the relative price of the primary input in terms of consumption, p, is assumed constant over time.

The firm is controlled by a workers' council and therefore follows policies intended to maximize its workers' utilities. Thus, because workers (households) are assumed to have perfect access to international capital markets, the workers' utilities are optimized by maximizing the present discounted value of the workers' income. More specifically, the representative firm's objective function is

$$\int_0^\infty \omega_t e^{-rt} dt. \tag{A-9}$$

Therefore the representative firm's problem is simply to maximize the present discounted value of its employees' incomes by choosing the path of the wage rate, subject to the flow-budget constraint (equation A-8) and initial real monetary balances, z_0. This is a linear optimization problem that has a "bang-bang" type of solution (Shell 1967). Let $z = z_\infty(p)$ be such that it satisfies the following equation:

$$\phi_n(z/p, p)/p = r. \tag{A-10}$$

Hence z_∞ is the level of real monetary balances such that the net marginal productivity of real monetary balances at the firm equals the international rate of interest. If $z_0 = z_\infty$, then it is optimal to set $z_t = z_0$ for all t (that is, z_∞ is the firm's optimal steady state). However, if $z_0 \neq z_\infty$, then ω will be set as large as possible (in absolute value) to reach z_∞ in the shortest period of time.

If no bounds are put on wages and if initial real monetary balances are larger than the steady-state optimum, then the firm should give an initial bonus to employees equivalent to the difference between the initial and optimum levels. Otherwise, if initial real money balances are less than the optimum level, the firm should order workers to bring to the firm's vault a stock of cash equivalent to the difference.

In such a liquidity crunch an initial transfer would be equivalent to a loan from workers to the firm yielding a rate of return larger than the international interest rate. The amount of the transfer is optimal, because, by equation A-10, the marginal return of liquidity at the firm equals the opportunity cost of loanable funds. In this fashion, therefore, the firm operates as if it had perfect access to international capital markets.

To capture credit segmentation in a simple manner, we will assume that the wage rate, ω, cannot fall below a well-defined minimum wage, $\omega > 0$ (the minimum wage is assumed to be smaller than long-run optimal output, that is, $\omega < \phi(z_\infty/p, p)$). Under this amendment the optimal policy for the firm remains the same as in the previous case (in which ω is subject to no bounds) if initial monetary balances at the firm, z_0, exceed the long-run optimum, z_∞. However,

if the firm starts on the liquidity-crunch region (that is, $z_0 < z_\infty$), then it is optimal to set the real wage, ω, at its minimum permissible level, $\underline{\omega}$, until real monetary balances at the firm, z_0, attain their optimal long-run level, z_∞. Thus, in symbols, along an optimal path we have

$$\text{(A-11)} \qquad \text{if } z_t < z_\infty, \text{ then } \omega_t = \underline{\omega}.$$

When firms are strapped for funds, they "borrow" from their own workers. This *is not an unusual phenomenon in socialist economies.*

Consider now the effect of a once-and-for-all devaluation of the currency, assuming that, beforehand, firms were at their long-run optimum liquidity, z_∞. A devaluation raises the domestic prices of inputs and outputs in the same proportion—hence, leaving p unchanged. Thus z_∞ remains the same, but now $z_0 < z_\infty$, that is, enterprises are subject to a liquidity crunch. By the above analysis, output falls, and firms react by lowering the real wage until liquidity returns to z_∞. Raising the relative price of the primary good, p, always leads to lower consumption output, but its effect on long-run optimal liquidity, z_∞, is ambiguous and will not be pursued here. During this process, output (and net output y) recovers monotonically to the original steady state. Consequently, the firm's optimal response satisfies all the conditions discussed in the first part of section III.

Government

To close the model, we assume that the government consumes nothing and returns to households its entire fiscal revenue (net of interest payments). Therefore, as in Calvo (1986), one can show that

$$\text{(A-12)} \qquad c = r(f_0 + \int_0^\infty y_t e^{-rt} dt)$$

where f is the country's overall stock of international bonds (which is the sum of the international bonds held by households and those held by government). Thus, recalling that y stands for output net of primary input costs, equation A-12 simply states that consumption equals "permanent" income. The latter, in turn, equals the return on total domestic assets: international bonds plus the present discounted value of net domestic output of consumption goods.

In the liquidity-crunch experiment in which all nominal prices are raised in the same proportion, we showed that y will initially fall and then steadily recover to its original steady-state value. Thus, by equation A-12, consumption falls on impact, but it does so by an amount that is less than the fall in initial output, y_0. Consequently, the trade balance initially deteriorates, as stated in the text.

Bank Credit

The model can easily be appended to allow for situations in which bank credit is available and official credit limits are binding. For example, let us denote total

credit and liquidity available to the representative firm by ℓ. Thus

(A-13) $$\ell_t = z_t + b_t$$

where b denotes bank credit to the representative firm. The cash-in-advance constraint (expression A-5) is now assumed to take the following form:

(A-14) $$\gamma p_t n_t \leq \ell_t.$$

Expression A-14 should thus be more properly referred to as a "liquidity-in-advance" constraint or, alternatively, as a "cash-and-credit-in-advance" constraint.

The interest rate on bank credit is denoted by r^b. Hence, if expression A-14 is binding, equation A-8 takes the following form:

(A-15) $$\dot{z}_t = \phi(\ell_t/p; p) - r^b(\ell_t - z_t) - \omega_t, \qquad \text{for } \ell_t \geq z_t.$$

In the previous version of the model we assumed that total liquidity, ℓ, was always equal to firm-owned liquidity, z. Thus, $z = \ell$, and the second term of the right side of equation A-15 drops out, yielding equation A-8, as expected. Otherwise total liquidity could exceed own liquidity, that is, $\ell > z$, but the firm would have to service the resulting bank debt, giving rise to extra cost equal to $r^b(\ell - z)$.

If bank credit is not binding, the firm is free to borrow as much as desired, and it is optimal for the firm to choose ℓ so as to maximize net-flow revenue as given by the right side of equation A-15. The first-order condition corresponding to such an optimum is (assuming $p = 1$ for notational simplicity)

(A-16) $$\partial\phi(\ell; 1)/\partial\ell = r^b.$$

From identity A-6 and equation A-16—and recalling that the liquidity-constraint (expression A-14) is assumed to be binding—raising the interest rate on bank loans, r^b, will result in output contraction and will induce early repayment of bank loans by enterprises (which is in line with developments in Poland during the first stages of the stabilization program).

Symmetrically, if the interest rate on bank loans, r^b, is lowered, output expands and so does the demand for bank credit by enterprises. For a sufficiently low r^b the credit ceiling (denoted by $\bar{b}$) will be binding. At that point the own-liquidity accumulation equation (A-15) becomes

(A-17) $$\dot{z}_t = \phi[(z_t + \bar{b})/p; p] - r^b\bar{b} - \omega_t.$$

One can show that during periods of credit rationing, that is, when the credit ceiling, $\bar{b}$, is binding, the economy displays the same qualitative behavior as in the first case discussed above, in which, implicitly, the credit ceiling, $\bar{b}$, was set equal to zero.

References

The word "processed" describes informally reproduced works that may not be com monly available through library systems.

Blanchard, Olivier, Rudiger Dornbusch, M. King, Paul Krugman, Richard Layard, and Lawrence Summers. 1991. *Reforms in Eastern Europe*. Cambridge, Mass.: MIT Press.

Calvo, Guillermo A. 1986. "Temporary Stabilization: Predetermined Exchange Rates." *Journal of Political Economy* 94 (December): 511–33.

———. 1989. "Incredible Reforms." In Guillermo A. Calvo, R. Findlay, P. Kouri, and J. Braga de Macedo, eds., *Debt, Stabilization, and Development*. Cambridge, Mass.: Basil Blackwell.

Calvo, Guillermo A., and Fabrizio Coricelli. 1991. "Stabilizing a Previously Centrally Planned Economy: Poland 1990." *Economic Policy*, forthcoming.

Cavallo, Domingo. 1977. "Stagflation Effects of Monetarist Stabilization Policies." PhD thesis. Harvard University, Cambridge, Mass. Processed.

Central Statistical Office. 1990. *Monthly Statistical Bulletin*. Warsaw.

Frydman, Roman, and Stanislaw Wellisz. 1991. "The Ownership-Control Structure and the Behavior of Polish Enterprises during the 1990 Reforms: Macroeconomic Measures and Microeconomic Response." In Vittorio Corbo, Fabrizio Coricelli, and Jan Bossak, eds., *Reforming Central and Eastern European Economies: Initial Results and Challenges*. A World Bank Symposium. Washington, D.C.

Gomulka, Stanislaw, and Jacek Rostowski. 1988. "An International Comparison of Material Intensity." *Journal of Comparative Economics* 12 (December): 475–501.

Kiguel, Miguel, and Nissan Liviatan. 1990. "The Business Cycle Associated with Exchange-Rate-Based Stabilizations." PRE Working Paper 513. World Bank, Country Economics Department. Washington, D.C. Processed.

Kornai, Janos. 1980. *Economics of Shortage*. Amsterdam: North Holland.

National Bank of Poland. 1990. *Monthly Bulletin*. Warsaw.

Osband, Kent. 1991. *Index Number Biases during Price Liberalization*. IMF Working Paper WP/91/76. Washington, D.C.: International Monetary Fund.

Shell, Karl. 1967. "Optimal Programs of Capital Accumulation for an Economy in Which There Is Exogenous Technical Change." *Essays in the Theory of Optimal Economic Growth*. Cambridge, Mass.: MIT Press.

Taylor, Lance. 1980. "IS-LM in the Tropics: Diagrammatics of the New Structuralist Macro Critique." In Sidney Weintraub and William R. Cline, eds., *Economic Stabilization in Developing Countries*. Washington, D.C.: Brookings Institution.

van Wijnbergen, Sweder. 1982. "Stagflationary Effects of Monetary Stabilization Policies: A Quantitative Analysis of South Korea." *Journal of Development Economics* 10: 133–69.

———. 1983. "Credit Policy, Inflation, and Growth in a Financially Repressed Economy." *Journal of Development Economics* 13: 44–65.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 91-108

Voluntary Choices in Concerted Deals: The Menu Approach to Debt Reduction in Developing Countries

Ishac Diwan and Kenneth Kletzer

This article examines the mechanics and attributes of concerted debt reduction agreements that offer creditors a choice between exit and relending options. The menu approach sets prices for different choices that implement a decentralized equilibrium. When banks can commit to choose from the menu and are not allowed to free ride, a menu can be designed that assures that the price paid for debt repurchased is equal to the marginal value of the debt claims. This can be achieved by taxing the gains that accrue to nonexiters with a request for new money. The equilibrium amount of debt reduction rises when the new money request is increased. The importance of banks' heterogeneity for menus to dominate simple concerted buybacks and the case in which debt reduction can be financed by loans from international financial institutions are discussed.

A debtor country would gain by repurchasing some of its outstanding debt if the price it pays is low enough. Similarly, there is some (high enough) price above which its creditors would gain by selling some of their debt claims. For a Pareto-improving debt reduction plan to exist, these two price ranges must overlap. When a Pareto-improving exchange is possible, finding a mechanism that implements such trade is an important remaining issue. In the case of country debt, market buybacks and concerted debt reduction are unlikely to do the job.

If debt is repurchased on the secondary market, then the price that must be paid is the equilibrium price of debt claims after debt reduction (Dooley 1988). This is because a market buyback by the debtor leads to a rise in the secondary market price. Since any lender can choose between selling and retaining debt claims, the opportunity cost for selling is the market price after the debt reduction. When lender participation in a buyback is voluntary, both the price paid for repurchased claims and the market price of remaining debt rise, and all creditors realize a net benefit at the expense of the debtor. The debtor is generally better off not participating in a market buyback if the funds to be used to

Ishac Diwan is with the International Economics Department of the World Bank, and Kenneth Kletzer is with the Economic Growth Center of the Department of Economics at Yale University.

repurchase debt are available to it for other purposes (such as for domestic investment or consumption).

Concerted agreements can be designed so that creditors commit to a coordinated debt reduction plan that is Pareto improving. Ideally, each creditor bank would sell a specific share of its claims at some price between the pre-debt reduction price and the expected post-debt reduction price. In practice, it may be difficult to bar individual banks from free-riding by not selling the agreed share. The problem of free-riding is exacerbated by the presence of heterogeneity across banks. If creditors differ in their relative valuations of country debt, a concerted buyback that does not discriminate between banks, and that at the same time hurts no bank, must occur at the reservation price of the bank with the highest valuation. With unobservable heterogeneity between creditors, the ability to discriminate between creditors in a concerted agreement will be limited. Offering a menu of choices could still achieve Pareto improvement by allowing creditors to self-select, with only those with low valuations selling out at a particular offer price.

Recently the menu approach has been used to reduce the external debt overhang of some highly indebted countries. An agreed upon menu is a contract, which may be partly implicit, establishing a future opportunity set for the lenders. The menu approach requires that lenders can commit to choose from among a restricted set of actions ex post. The possible means for enforcing a menu include creditor country legal and political institutions and agreements between creditor nations. By agreeing to the contract, lenders face the possibility that they will be penalized for reneging ex post by refusing to allocate their existing debt claims across a portfolio using the menu options. While private creditors restrict their options ex post, they may be able to raise their net worth in negotiations. That is, lenders can increase the value of their loan portfolios ex ante by agreeing to choose from a menu of restricted options ex post.

I. The Menu Approach

The menu approach to debt reduction combines characteristics of voluntary market buybacks and concerted debt reduction agreements. The menu approach thus retains the advantages, but not the inconveniences, of the other two mechanisms. The options on the menu and their relative pricing are negotiated first; in a second round each creditor freely chooses a preferred option. Overall, the discrimination allowed by the menu leads to cheaper debt reduction (Diwan and Spiegle 1990).

This article analyzes the mechanics of simple debt-reducing menus that include exit and relending options. Several types of exit options can be offered to fit the regulatory and tax characteristics of different types of banks. But often some banks would require too high a price to exit irrespective of the exit vehicle. Therefore a nonexit option is included in the menu in order to achieve agreement on a low exit price. The creditors that do not choose to exit get a capital

gain following the reduction in debt outstanding (because the price of debt increases when the debt stock decreases). Unless this gain is taxed, the nonexit option would dominate the exit option, and no exit would occur freely. This problem can be mitigated by imposing a capital gains tax on those banks that choose the nonexit option.

While explicit taxation of creditors' gains may be politically or legally infeasible, there may be alternative schemes that effectively tax these windfall gains to the nonexiting debt-holders. Such a "tax" can be set by requiring lenders to extend new loans in proportion to the debt they retain. A commitment to provide new loans in proportion to retained debt creates a wedge between the price paid to repurchase debt and the secondary price of outstanding debt after the stock of debt has been reduced. In recent menu agreements, the options have included some form of buybacks and of new loans. When the new debt claims trade at a discount, a portion of any new loan represents such a tax. That is, if a dollar is lent and the secondary price is 60 cents, then the new debt claim is worth 60 cents and the rest is a tax with the revenue accruing to the debtor.

The menu approach does not require assignment of particular choices from the menu to each lender. By offering a set of prices for different options, the menu leads to a decentralized equilibrium level of debt reduction ex post. There is a unique level of equilibrium debt reduction that corresponds to a given set of options because, in equilibrium, some banks would have to be indifferent between the two options. The equilibrium amount of debt reduction is affected by the terms of the menu. When the new money call is set higher, more banks will prefer the exit option. The ex-post debt price will then increase, making the new money option more desirable.

For the menu approach to debt reduction to create gains over and above those available with simple concerted buybacks, the offered options must add to existing market trading opportunities. In general the use of the menu approach allows for more gains when the secondary market for debt is inefficient, either because of incompleteness, distortionary public policies, or large transaction costs. When the cost of debt reduction is (partly) financed by new international loans, the results depend on whether the new loans share in the same seniority class as the old loans.

II. A Debt Valuation Model and the Menu Approach

Sovereign borrowers choose to service external debt obligations because it is in their enlightened self-interest to do so. Under sovereign immunity, creditors do not have recourse to an international court to assure settlement of debts through a direct lien on the country's assets as they would in the case of an insolvent domestic client. The threat of the imposition of sanctions in the event of nonpayment or insufficient payment provides sovereigns with the incentive to make debt payments. In addition to denials of future official aid flows, the penalties for default include restrictions on future trading opportunities on inter-

national markets for the debtor. These include the disruption of commodity trade, suspension of trade preferences, and reduced access to international financial markets. Because the repayments made by a sovereign debtor are linked to the expected present value of the social cost of sanctions, they need not be strictly proportional to the total outstanding contractual debt burden. Moreover, as the probability of sanctions being exercised (or being used by creditors as a threat to extract higher net repayments) increases, debtor countries can reduce the effectiveness of sanctions, for example, by underinvesting or by shifting resources to the nontradable and to the import-competing sectors of the economy.

A simple model summarizes the various arguments made by the recent debt literature on the effect of market buybacks and contrasts those with the effects of menu-driven debt reduction. The model—a system of equations—can be used to determine how much debt reduction can be achieved with a given amount of resources or how much resources would be required to reduce debt by a given amount.

A Simple Model of Debt Valuation

We assume that the present value of expected debt service payments is an increasing and concave function of the nominal debt burden. In other words, the value of debt is a smooth and concave function ($f(D)$ in figure 1) of the amount of outstanding total debt claims, D. In this analysis we consider only the case in which the marginal price of debt is positive: the debtor is on the increasing part of its debt value curve. But if the debt value curve is first upward, then downward sloping (as in the lower half of figure 1), the arguments below are unchanged. The debt value curve includes the effect that the debt burden has on the investment undertaken by the debtor; if debt reduction raises investment, the increase in the expected present value of repayments made on outstanding debt is included in the debt value curve.

If creditors sell a portion of the initial debt they hold, the secondary market price of debt will rise from p_0 in equation 1 to p_1 in equation 2:

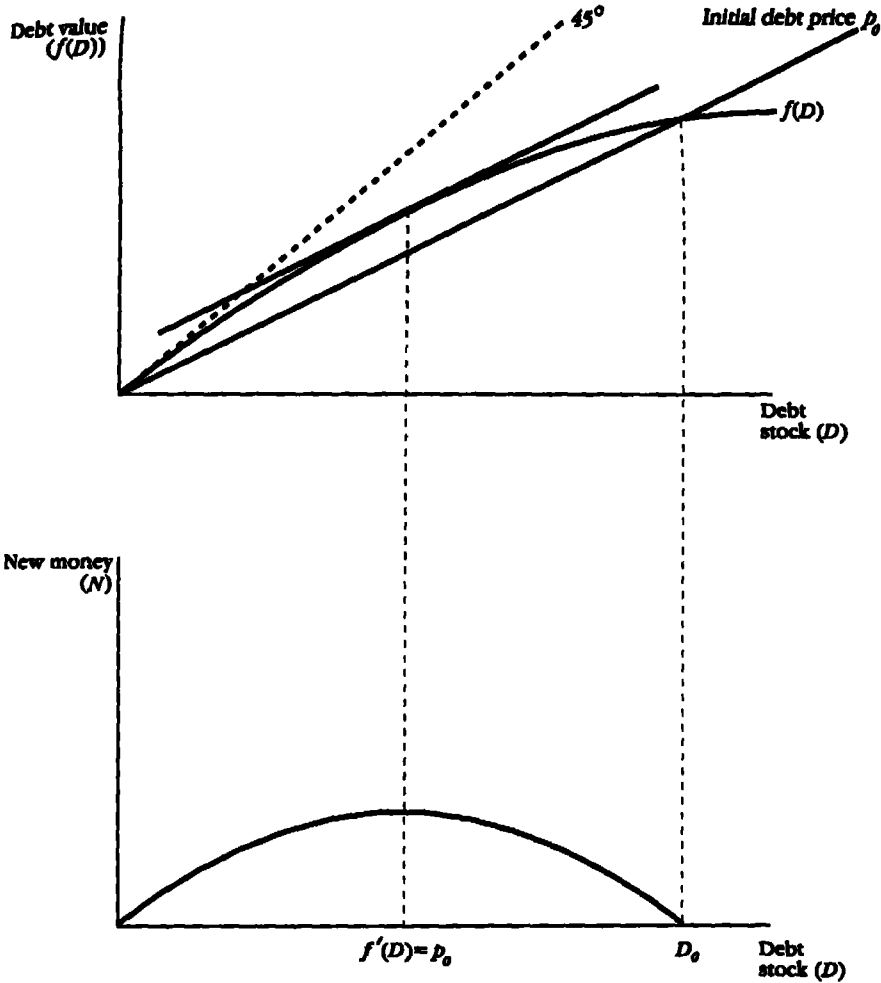
$$p_0 = f(D_0)/D_0 \tag{1}$$

$$p_1 = f(D_0 - \alpha/p_1))/[D_0 - (\alpha/p_1)] \tag{2}$$

where p_0 is the initial debt price, D_0 is the initial stock of debt, p_1 is the new secondary market price of the reduced debt, and α is the amount of funds used to purchase debt. Because the debt value function is concave, the new secondary market price of debt exceeds the initial debt price for all positive values of debt reduction. The ex-post value of creditor assets ($f(D_1)$) is:

$$f(D_1) = p_1(D_0 - \alpha/p_1) + \alpha = p_1 D_0. \tag{3}$$

The ex-post value of creditor assets is greater than the original value of debt claims because the new debt price is greater than the initial debt price. The capital gain (γ) realized by the creditors retaining claims is

Figure 1. *The Debt Value Function and Equilibrium New Money*

$$\gamma = (p_1 - p_0)(D_0 - \alpha/p_1) > 0. \quad (4)$$

All creditors realize a capital gain when the price of debt increases, whether or not they have sold a portion of their claims.

Automatic Equilibrating Mechanisms in Simple Menus

When contracts that bind the actions of private creditors are feasible and credible, the conversion of the status quo debt claims to a package of assets that has expected present value at least as great can be negotiated. It is assumed that

enforcement of an agreement by private creditors with their home governments and international financial institutions is possible, but that creditors do not need to accept a contract that reduces the present value of their assets. Once an agreement is reached, lenders face penalties if they refuse to perform according to the contract (in the recent Mexico deal, for example, creditors that did not participate were effectively made junior). If enforcement of a contract is possible only for a subset of the creditors, then a debt reduction menu can still be negotiated. However, the nonparticipants free ride, thus realizing a net increase in the value of their portfolios. The participating creditors can increase the value of their assets despite providing pecuniary external benefits to the nonparticipants. For simplicity, it is assumed that the entire stock of outstanding claims is represented in the negotiations.

Suppose that the creditor banks and the debtor country have agreed on a simple menu of options represented by (p,n), in which, for each dollar of claim they hold, banks can choose to either exit at price p or to reschedule the loan and relend n dollars in addition (where n is a percentage of the debt they retain). With no loss of generality, we assume that new and old loans have equivalent terms. Given a simple menu, can we tell how much debt reduction and new money will actually be achieved? For simplicity, the equilibrium analysis below is developed for the case in which all banks are similar. The more general case with heterogeneous banks is discussed in section IV and more fully in Diwan and Spiegle (1990).

After the deal is completed, debt prices are expected to be higher. A bank that relends a portion of its debt will have its old (rescheduled) claim revalued and will realize a capital gain as indicated in equation 4. However, its new claim will be valued at the secondary market price, implying a capital loss (equal to $1 - p_1$ per unit of new debt). Thus the opportunity cost for holding a unit of debt back from repurchase at the price agreed in the menu depends on the increased secondary market price and the amount of new money loaned. The opportunity cost, C, of holding a unit of debt back from repurchase at the increased secondary market price is given by the following:

$$C = p_1 (1 + n) - n. \tag{5}$$

The opportunity cost of holding a unit of debt is equal to the exit price when the ex-post debt price satisfies the no-arbitrage condition in equation 6:

$$p_1 = (p + n) / (1 + n). \tag{6}$$

When the ex-post secondary market price of debt is expected to exceed the right side of equation 6, the new money option is preferred to the exit option. Banks would rather hold onto their debt than sell at an exit price of p because the capital gain on old claims is greater than the capital loss on agreed (by the menu) new lending. Thus less debt will be sold and more new money offered, resulting in less debt reduction. This leads to an increase in debt stock and thus to a reduction in the ex-post price. Because creditors are price takers when they optimize ex post and because the expected present value of the debt function is

strictly concave, the solution to portfolio value maximization by creditors is unique. Equation 6 must hold in equilibrium.

When "sufficient" funds are available for debt reduction, competition among banks will ensure a unique equilibrium. The menu approach does not require that the ex-post choice of each creditor from the menu be negotiated and specified or even that aggregate constraints be placed on the total amounts of debt reduction and new money. Prices, that is, the offer of the initial debt price for debt bought back, and any ratio of new loans to remaining debt claims required to obtain that price are sufficient.

For a given amount of resources for debt reduction, we can calculate by how much debt stocks will be reduced. Given the initial debt stock, debt value function, and menu (exit price and relending ratio agreed on), equations 6 to 9 can be solved for new debt stock, amount of debt reduction, and new lending.

$$D_1 = D_0 - B + N \tag{7}$$

$$p_i = f(D_i) / D_i, \qquad i = 0,1 \tag{8}$$

$$n = N / (D_1 - N) \tag{9}$$

where D_i denotes debt stocks (subscript $i = 0$ indicates the level of the variables before the debt agreement and subscript $i = 1$ indicates the level after the deal is completed), B denotes the amount of reduction of debt stocks, and N denotes new lending. Lenders choose between the menu's two options (selling or relending) in a manner that maximizes the value of their assets subject to the terms of the menu.

What is remarkable is that any menu will produce an equilibrium in which all the banks, whether they exit or relend, retain a payoff exactly equal to the offered exit price. Also, when the menu approach is used to implement a debt reduction, increases in debt prices following the debt reduction are in the interest of the debtor since it receives a compensating transfer from the nonexiting creditors. This is in contrast with the pure market buyback case in which the opposite holds (Bulow and Rogoff 1988). Thus any menu effectively allows the debtor to capture a share of the entire efficiency gains generated by debt reduction. The actual gains that accrue to the creditors depend on the level of the agreed exit price. The closer the exit price is to the ex-ante (initial debt) price, the less the banks gain. But the closer the exit price is to the ex-post (increased secondary market) price, the larger is their gain. There are two limiting cases of interest: when the exit price is exactly set to the ex-post price (as happens with pure market buybacks) all efficiency gains go to the creditors and when it is set to the initial debt price, in which case the debtor captures all the efficiency gain.

III. Characteristics of the Menu Approach to Debt Reduction

This section looks at comparative statics associated with changes in the terms of the menu exit and relending options and discusses the costs associated with the debt deal.

Comparative Statics

With the menu approach the equilibrium amount of debt reduction varies depending on the agreed levels of the exit price and the relending ratio. From equations 7 to 9:

$$f(D_1) / D_1 = (p + n) / (1 + n). \quad (10)$$

Differentiating equation 10 with respect to the exit price, p, and the relending ratio, n, and rearranging yields:

$$dD_1 / dp_0 = -1 / \{D_1 (1 + n) [p_1 - f'(D_1)]\} \quad (11)$$

$$dD_1 / dn = -[D_1 /(1 + n)] \{(1 - p_1) / [p_1 - f'(D_1)]\}. \quad (12)$$

Equations 11 and 12 are both negative when the debtor is on the increasing part of its debt value curve.

Thus, when the relending ratio is held constant and the exit price is increased, more debt reduction is achieved in equilibrium. This is because as the initial exit price rises, exit becomes more desirable. More debt reduction is then needed in order to increase the secondary market price sufficiently so that the relending option becomes as valuable as the exit option.

Holding the initial debt price constant and increasing the relending ratio increases the equilibrium amount of debt reduction achieved. As the relending ratio is increased, the exit option becomes more desirable than the relending option. But, in equilibrium, both options must be equally desirable. As a result a larger debt reduction will be achieved in order to raise further the ex-post debt price and increase the attractiveness of the relending option.

The Cost of the Deal

In the simple model, given knowledge of the debt value function, the new relending ratio and exit price can be determined so the deal costs exactly the amount of funds available to repurchase debt, α. If there is uncertainty about the debt value function, then any particular menu will lead to a random amount of debt reduction (and consequent demand for buybacks). Given an exit price, there need not be a windfall gain to creditors, but the amount of resources required to "back the deal" will be uncertain. If the amount of relending turns out to be too large, only a portion of the available resources will be used, but creditors will remain at the payoff level implied by the exit price. However, if the amount of relending turns out to be too low, the available resources will not be sufficient to buy back all the tendered debt. If this occurs, a distribution mechanism will be needed to allocate the scarce buyback resources. For example, buybacks can be distributed on a pro-rata basis. If all banks are similar, they all end up with a similar mix of cash and new loans, per unit of old debt, and this payoff would be lower than that implied by the exit price. Because of this possibility creditors will initially insist on a higher exit price to protect themselves when the debt value function is uncertain. Since relending may not end up

as desirable as exiting, the exit price must be large so as to produce a portfolio of a given value.

An underfunded deal is not in the interest of the debtor country. An adequately funded menu will allow the debtor to discriminate more fully between different types of creditor banks. As discussed below, banks are likely to be heterogeneous with respect to their valuation of country debt. Since equilibrium deals allow banks to self-select from the menu the options that they value most while ad hoc distribution mechanisms do not, equilibrium deals with adequate funding allow the debtor to reap additional gains (ex ante) over deals in which options are distributed in some ad-hoc way.

IV. The Menu Approach with Heterogeneous Banks

To simplify the equilibrium analysis of menu choices, it was assumed that all banks are similar. Removal of the assumption of homogeneous banks does not affect the equilibrium and comparative statics results in any important qualitative way. If banks value country debt differently, the banks may be ranked by their valuation. A simple menu divides the banks into two groups: low valuators, which value the exit option more than the new money option, and high valuators, which place greater value on the relending option. Separating these two groups is a bank that is indifferent between the two options. For this marginal bank, a no-arbitrage condition applies (equation 6) such that the exit price given by the menu is equal to the value of the relending option. The comparative statics are then similar to those in the homogeneous case. In particular, when the new money call increases, the exit option becomes more valuable to all banks, reducing the ex-post value of debt to all banks. As a result, the marginal bank shifts toward banks with higher valuation, and the equilibrium amount of debt reduction increases (see Diwan and Spiegle 1990 for details).

Because it values the two options equally, the marginal bank retains no surplus from the deal. Banks on both sides retain a surplus that increases with the distance from the marginal bank. Thus, when banks are not homogeneous, a single "new money" tax is not sufficient to ensure that the "surplus" of every lender is the same. But the menu dominates simple concerted debt reduction deals because exit would not occur in a concerted agreement at any price below the valuation of the bank with the highest valuation. More gains can possibly be achieved for the debtor by extending the menu and achieving finer discrimination. For example, several new money and exit options could be offered that differ in the timing of the payments made by the banks.

Menus would not be needed if banks were homogeneous. A buyback price could be agreed on for which all banks would agree to sell a certain proportion of their portfolio at this price. A pure concerted approach, rather than an approach that combines concerted with market features, would be sufficient. Banks' heterogeneity is the principal reason why the menu approach to debt reduction dominates a simple concerted approach.

If the secondary market for sovereign debt is efficient, the marginal value of a debt claim should be the same for all participating banks since the market intermediates away banks' heterogeneity. Banks would then be effectively homogeneous. Heterogeneity in valuation occurs only if the market is not fully efficient in intermediating the valuations of different banks. This can be the case if asset markets are incomplete. If lenders are unable to completely diversify against the risk of country debt using all available assets, then the change in the surplus caused by an overall debt reduction can vary across lenders even though the value of a marginal claim is the same. Reduction of the debt outstanding changes the distribution of returns to holding a part of the country's debt. The expected present value of the stock of debt held by each lender will not change by the same percentage, in general. This means that a single capital gains tax rate, imposed using the new money option, is insufficient to tax away all the potential gains to heterogeneous lenders.

Heterogeneity may also occur because there are distortionary public policies in place (such as subsidized deposit insurance) or there are some other types of transaction costs. Distortionary public policy and transactions costs create a wedge between the price at which debt is sold and the net payout that accrues to the seller. As a result, debt is not transacted in the secondary market when differences in valuation are smaller than this wedge. It can be shown that this wedge increases as the quality of the debt claims deteriorates (see Demirguc-Kunt and Diwan 1990).

If the secondary market for debt is inefficient, banks can differ for a number of reasons in spite of the possibility of trading. For example, a commercial bank may have other business interests in the debtor country, thus creating an extra benefit to lending (Sachs 1989). Other reasons for heterogeneous banks include differences in their expectations (Williamson 1988), differences in their nationality and thus in their tax and regulatory environment (Bouchet and Hay 1989), alternative business opportunities, and bank size in the presence of fixed costs associated with recontracting and monitoring. These factors give rise to a secondary market in the first place, but an inefficient market will not eliminate the effects of heterogeneity. Heterogeneity across lenders also implies that the choices the banks make from a menu will depend on their own characteristics. This issue is explored in depth in Demirguc-Kunt and Diwan (1990), using the experience of the Brazilian 1988 debt deal. They show that 80 percent of bank choices between exit and relending options can be explained by measures of financial strength, exposure, nationality, and long-term interest in the country.

V. The Menu Approach with Debt Reduction Financed by International Financial Institutions

If funds are supplied by international financial institutions seeking to benefit the debtor, then a package that effectively taxes the capital gains of lenders will reduce the net transfer to the commercial creditors. Commercial lenders can

agree to choose between selling debt back to the debtor and holding claims with the obligation to provide new loans. A menu can be designed to ensure that lenders are indifferent between the two options. The status quo payoff can be determined if the relationship between the present value of debt and the face value of debt is known.

In general, the new loans forthcoming under the relending option can be used by the debtor country either for domestic absorption or for further debt repurchases. If they are used for buybacks, the funds available for buybacks are increased above the contribution of the international financial institutions. The debt is then reduced further, and the relending ratio should be higher since the secondary market price rises further.

After the debt reduction program, the remaining debt is given by:

$$D_1 = D_0 - [(F + N - \mu)/p] + N \tag{13}$$

where F is the amount of funds supplied by the international financial institutions, N is the total of new loans made by the remaining creditors, μ is the part of the new financing $(N + F)$ kept by the debtor country to finance domestic absorption, and p is the price paid in the buyback. It is assumed that the net resource transfer to the country, μ, is a monotone increasing function of the funds supplied by the international financial institutions $(N + F)$ and the new loans made by creditors. The amount of new funds used to repurchase debt is equal to the funds supplied by the international financial institutions plus new loans by creditors minus the funds used to finance domestic absorption $(N + F - \mu)$.

In equilibrium the following no-arbitrage condition must now hold:

$$N + f(D_0) = (F + N - \mu) + f(D_1). \tag{14}$$

The left side of equation 14 is the value of the creditors' assets before the buyback. The right side expresses the value of their assets after the operation has been completed. In addition to the value of remaining debt, the creditors receive a net amount $(F + N - \mu)$. When the funds supplied by international financial institutions, F, exceed the net resource transfer to the country, μ, the private lenders gain. In this case, if required, the private lenders are willing to pay a positive price for debt reduction. The amount of new money necessary for creditors to retain no net increase in the value of their portfolios is positive. There is a positive equilibrium value of new loans by creditors, given the level of funds supplied by international financial institutions, such that equation 14 holds whenever a part of the funds supplied is spent on buybacks.

The new secondary market price, which is a function of the amount of funds supplied by international financial institutions and the net resource transfer to the country, is higher than the initial debt price whenever the equilibrium level of new loans made by creditors exceeds zero. This follows from the concavity of the debt value function.

When debt reduction creates no change in the value of creditors' assets, the

implicit tax rate imposed on remaining debt is the difference between the new secondary market price of debt and the initial debt price. The capital gains realized by remaining debt holders are taxed 100 percent. The capital gain due to debt reduction is $\gamma = [D_0 - (F + N - \mu)/p_0] (p_1 - p_0)$, and the total tax revenue, which accrues to the debtor, is $\lambda = N(1 - p_1)$. The equilibrium level of new lending by creditors is obtained when the capital gain is equal to the tax revenue. This is equivalent to solving equation 14. The amount of new money required per unit of old debt held back from repurchase is

$$(15) \qquad n = N / (D_1 - N) = (p_1 - p_0) / (1 - p_1)$$

which is equivalent to equation 6. Since the buyback occurs at the ex-ante price, the exiting creditors also break even.

The effect of the funds supplied by international financial institutions on the equilibrium level of new loans by creditors and the new equilibrium level of debt stocks can be calculated using equations 13 and 14:

$$(16) \qquad dD_1/dF = -(1 - \mu')(1 - p_0) / [(1 - p_0)f' + (p_0 - f')\mu'] \text{ and}$$

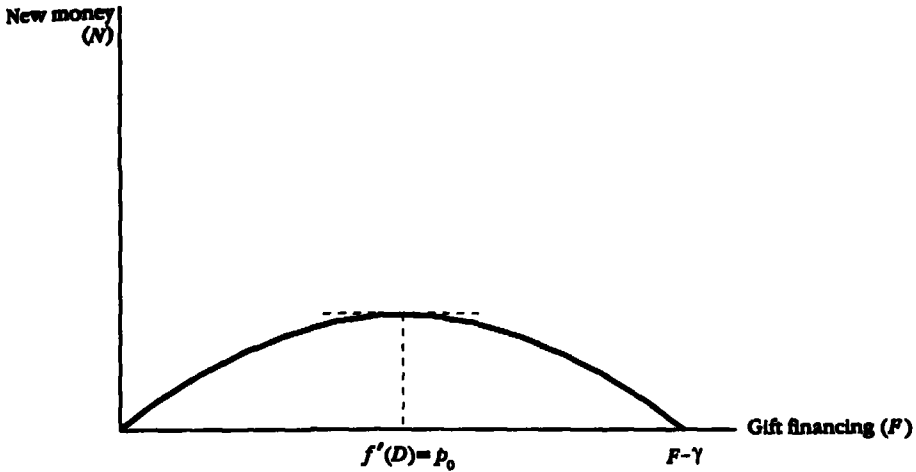
$$(17) \qquad dN/dF = (p_0 - f')(1 - \mu') / [(1 - p_0)f' + (p_0 - f')\mu']$$

where $f' = f'(D_1)$ and $\mu' = \mu'(F,N)$. When the marginal net resource transfer to the country is less than one, a rise in the funds supplied by international financial institutions increases both the equilibrium level of new loans by creditors and the new equilibrium level of debt stocks until the equilibrium marginal value of debt equals the initial debt price ($f'(D) = p$).

Further debt reduction beyond this point through increases in financing provided by international financial institutions decreases the size of the net resource transfer. As long as the marginal value of debt is below the buyback price (initial debt price), remaining debtholders realize an increasing capital gain from further repurchases. Therefore, given the constraint that creditors are at least as well off with the program as the status quo, the compensating concession in new money that can be demanded rises with debt reduction. However, when debt is reduced far enough so that the marginal value of debt is greater than the initial debt price ($f'(D) > p_0$), the implicit tax needed to assure that creditors are indifferent between selling and holding debt declines with further reductions in the debt. If the proportion of available funds used to repurchase debt can change, then an increase in $\mu(F + N)$ for every value of $(F + N)$ acts as a decrease in F.

Thus, until debt is reduced to the point that the marginal value of secondary debt equals the buyback price, the debtor gains both net resource transfers and debt reduction from further buybacks. If the objectives of the donor place weight only on the debtor's benefits (excluding any other objectives, such as the cost of funds), then the optimal amount of remaining debt for the donor is less than or equal to the solution to setting the marginal value of debt equal to the initial debt price ($f'(D) = p$).

Figure 1 depicts the debt value curve and initial value of debt in the top panel. The relationship between new loans made by creditors and the ex-post equilib-

Figure 2. ***Equilibrium New Money and Gift Financing***

rium level of debt is portrayed in the lower panel for solutions to equations 13 and 14. The maximum amount of new money contributed by creditors is attained when the marginal value of debt equals the initial debt price. Figure 2 illustrates the relationship between the equilibrium level of new loans by creditors and the amount of gift financing by international financial institutions. Equilibrium lending rises as financing increases until the marginal value of debt equals the initial debt price and then declines toward zero. The debt stock is zero when the gift is large enough to buy back the entire initial debt stock at the initial debt price.

When debt is repurchased with a loan, rather than a gift, from official sources, seniority considerations also matter. When all debt, regardless of the creditor, has equal priority in repayment, the results developed above occur. The only difference is that when part of the initial transfer is a loan, less new money is forthcoming than in the pure gift case because the new loan reduces the ex-post debt price. Consequently, less debt reduction can be achieved for a given transfer. However, when the initial transfer takes the form of senior debt, it impairs the value of existing debt claims since junior debt is now serviced only after the senior loan is repaid. In that case, despite the gains that are due to debt reduction, the value of private creditors' assets is reduced (see the appendix for a treatment of financing by international financial institutions as junior and senior debt).

It is probable that loans from international financial institutions are senior to

commercial loans, but that they also create efficiency gains due to conditionality (that is, they lead to an upward shift in the debt value curve). The net effect of those loans on the payoffs of commercial creditors can then be positive. The menu approach can help ensure that those net benefits are taxed so that the debtor country can retain a larger share of the efficiency gains due to conditionality.

VI. Concluding Remarks

When sovereign debt trades at a discount on secondary markets, a market buyback leads to an increase in the secondary market price. The wealth of private creditors increases because part of the funds used in the buyback is a transfer payment to them. When banks are homogeneous, the inclusion of exit and relending options in a menu from which private creditors freely choose can eliminate the wealth transfers due to the debt reduction. It is sufficient to set the buyback price equal to the ex-ante status quo price. Any new money call will do the job. In equilibrium creditors will provide enough new money so that the ex-post price leaves them indifferent to the exit option. An increase in the new money call reduces the cost of the menu and the extent of debt reduction achieved. Commitment by the banks to a menu including this new money option assures that debt repurchased is bought back at its marginal price instead of at its average price.

The menu approach does not require assignments of particular choices from the menu to each lender. Rather, it implements debt reduction through a price system, allowing different creditors to select different portfolios in equilibrium from a common set of options. With heterogeneous banks, some transfer of resources will occur when participation in the debt reduction plan is voluntary, and the buyback price will generally need to be above the pre-buyback price.

Appendix. The Menu Approach with Loans from the International Financial Institutions

Debt may be repurchased with a loan from official sources interested in providing debt reduction and new liquidity. In the first case considered, all debt, regardless of the creditor, has equal priority in repayment. That is, a government or international financial institution does not possess legal seniority privileges in relation to private creditors; this institutional assumption is ad hoc. The second case considered explores the effect that seniority privileges would have if the debt of international financial institutions was senior to private debt. For simplicity, only the special case in which the entire new money tax is used for domestic absorption is analyzed.

A loan can be made by the third party in several ways, one of which is equivalent to providing private new money. The loan amount can be provided in

exchange for a debt obligation of equal face value. Therefore, the expected present value of the acquired debt is less than the loan amount, and the agency is part donor. Another extreme is that the nominal debt obligation exceeds the loan amount by the discount on debt, so that the expected present value of the new debt obligation is equal to the amount paid (the loan amount).

The Case of Equal Priority in Payment

In this case if the expected present value of the new debt obligation is equal to the loan amount, the additional debt bought at the secondary market discount cannot reduce debt unless private creditors are forced to accept losses. This is because the reduction in private debt is made up one for one with new debt. Any new money tax imposed on private creditors results in an ex-ante appropriation of their wealth.

If the expected present value of the new debt obligation is less than the new loan amount, the new money loan can yield both debt reduction and more private loans. If the new loans are entirely used to buy back debt, then private creditors are indifferent between the status quo and equilibrium debt reduction under the menu if equations A-1 and A-2 hold:

(A-1) $$f(D_0) + N = p_1 [D_1 - L] + L + N$$

(A-2) $$D_1 = D_0 + N + L - (L + N)/p$$

and where $p_1 = f(D_1)/D_1$, where D denotes debt stock, N denotes new lending, p denotes the price of debt, L denotes the loan, and subscripts 0 and 1 indicate initial and after the deal, respectively. The right side of equation A-1 is the value of the assets held by all private creditors ex post. It also expresses that the ex-ante price of debt equals the ex-post price minus the tax imposed through new loan commitments. When the buyback price is set equal to the initial debt price, equation A-1 (for concave debt value functions) can be solved to obtain the positive equilibrium level of new lending. Creditors will be indifferent between selling debt claims at the ex-ante price and holding them with the commitment to provide a proportionate share of the equilibrium new money.

The effect of an increase in the loan amount on the extent of debt reduction and net resource transfer to the debtor can be calculated from equations A-1 and A-2:

(A-3) $$dN/dL = [p_0 (1 - p_1) / (1 - p_0)A] - 1$$

(A-4) $$dD_1/dL = -(1 - p_1)/A < 0$$

where A is given by

(A-5) $$A = (1 - L/D_1)f'(D_1) + (L/D_1)p_1.$$

Comparison with the gift case in section V reveals that when part of the initial transfer is a loan, less new money is forthcoming than in the pure gift case

because the new loans reduce the ex-post debt price. Consequently, less debt reduction can be achieved for a given transfer.

The Case of Senior Loans by the International Financial Institutions

In this case the new loan provided by an international financial institution impairs the value of existing debt claims. This is because after a senior international financial institution loan is made, the value of private debt equals the present value of debt claims only when they are serviced after the senior loan is repaid.

If the senior loan is used to repurchase debt and no new money is required, then the buyback price, p_1, is given by:

$$p_1 = [f(D + F - B) - f(L)] / (D - B) \quad \text{(A-6)}$$

where F is the amount of new funds provided by the international financial institution and B is the amount of debt repurchased. The amount of the loan used for buybacks is the buyback price times the amount of debt repurchased. If all of the new funds are used for buybacks, then the following equilibrium condition holds:

$$L/B = [f(D + F - B) - f(L)] / (D - B) \quad \text{(A-7)}$$

from which the equilibrium buyback price can be derived. Since the debt value function is concave, this has a solution for $F < B$, if $0 < F < D$.

The top part of figure A-1 depicts an equilibrium price when a buyback is financed by a senior loan from an international financial institution. The amount of the loan used to repurchase B units of debt is shown on the right side of the figure as p_1B. For the values of the ex-post price and the amount of debt repurchased shown, the demand for buybacks does not use up all of the loan. When all of the loan is used to repurchase debt, the equilibrium values of debt repurchased and the ex-post price are larger, so that p_1B equals the loan amount. The bottom part of figure A-1 shows the effect of dilution: p_0 is the price of private debt before the senior loan is made, and p_1 is the ex-post equilibrium price.

Comparison of the ex-post equilibrium price paid in a buyback to the ex-ante price (before the senior loan is made) shows that the value of private creditors' assets is reduced, and the present value of all expected debt repayments falls. Letting F/B equal the ex-post price, p, equation A-7 implies the solution for the ex-post price:

$$p = [f(D + F - L/p) - f(L)] / (D - F/p). \quad \text{(A-8)}$$

As the ex-post price times the debt stock approaches the loan amount (all private debt is repurchased), the right side of equation A-8 remains less than unity, in contrast to the case of equal seniority. Therefore a solution to equation A-8 exists for an ex-post price and loan amount such that all the private debt is repurchased and the ex-post debt burden is smaller than the initial debt stock.

Figure A-1. ***Debt Reduction Financed with a Senior Loan***

This reduction in the expected present value of debt repayments is financed by the dilution of the private creditors' debt claims caused by the senior loan. The difference, $f(D) - f(L)$, is the net transfer from private creditors to the debtor; this is a transfer of future rather than current resources.

To further illustrate the dilution effect of an increase in senior debt, suppose that private creditors are offered the ex-ante secondary market price, p_0, in the buyback (say for political reasons). From equation A-8, if the repurchase price is set to $p_0 = f(D)/D_0$ on the right side, then the solution for the ex-post secondary market price of private debt (appearing on the left side) is less than p_0. Remain-

ing private creditors suffer a capital loss and prefer to sell back debt. An equilibrium buyback price such that creditors are indifferent at the margin between selling and retaining debt is lower than the initial price.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Bulow, Jeremy, and Kenneth Rogoff. 1988. "The Buyback Boondoggle." *Brookings Papers on Economic Activity* 2.

Bouchet, Michel, and Jonathan Hay. 1989. "The Rise of the Market-Based Menu Approach and Its Limitations." In Ishrat Husain and Ishac Diwan, eds., *Dealing with the Debt Crisis*. A World Bank Symposium. Washington, D.C.: World Bank.

Demirguc-Kunt, Asli, and Ishac Diwan. 1990. "The Menu Approach to Developing Country External Debt: An Analysis of Commercial Banks' Choice Behavior." PRE Working Paper 530. World Bank, International Economics Department, Washington, D.C. Processed.

Diwan, Ishac, and Mark Spiegle. 1990. "Menu-Driven Debt Reduction Schemes with Heterogeneous Creditors." World Bank, International Economics Department, Washington, D.C. Processed.

Dooley, Michael. 1988. "Buybacks and the Market Valuation of External Debt." In Jacob Frenkel and others, eds., *Analytical Issues in Debt*. Washington, D.C.: International Monetary Fund.

Sachs, Jeffrey. 1989. "Efficient Debt Reductions." In Ishrat Husain and Ishac Diwan, eds., *Dealing with the Debt Crisis*. A World Bank Symposium. Washington, D.C.: World Bank.

Williamson, John. 1988. *Voluntary Approaches to Debt Relief*. Paper 25. Washington, D.C.: Institute for International Economics.

WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 109-124

Does Undernutrition Respond to Incomes and Prices? Dominance Tests for Indonesia

Martin Ravallion

ecent evidence suggests that food energy intakes of the poor respond less to income han was once thought. However, it is not intake alone that is of concern, but under-utrition. Two facts confound assessments of how undernutrition responds to incomes nd prices: individual nutrient requirements vary in a generally unobserved way, and ntakes are observed with error. By modeling observed intake distributions economet-ically, straightforward stochastic dominance tests can permit robust qualitative infer-nces about such responses. An application to Indonesia in the mid-1980s indicates that egional distributions of food energy intake are influenced by average income levels, ntraregional inequalities, and local prices of staple food grains—all of which have nambiguous effects on undernutrition. The results suggest that any adverse effects on nequality of a growth process would need to be large to outweigh the desirable effect n undernutrition. In addition higher food staple prices still have adverse effects on ndernutrition after allowing for their likely positive effects on rural incomes.

: attainment of adequate nutritional levels is an important criterion in eval-ing the success of development policies. However, such evaluations have n been hampered by both conceptual and technical problems associated with v undernutrition is measured. Most important among these is that the nutri-ial requirements for good health vary across individuals and over time in erally unknown ways, and nutrition intakes are typically measured with or. In light of these problems, what can we say about how the instruments of elopment policies influence the extent of undernutrition?

)evelopment policies have often emphasized the role of economic instru-its, particularly incomes, in reducing undernutrition. However, recent isehold-level studies raise doubts about the effectiveness of these instruments

rtin Ravallion is in the Welfare and Human Resources Division of the Population and Human urces Department at the World Bank. This article is a product of a research project supported by the d Bank's Research Committee. The author is grateful to Monika Huppi and Alok Bhargava for their with the data and estimation and to Indonesia's Central Bureau of Statistics for providing the data of the 1984 and 1987 National Socioeconomic Surveys. Helpful comments on the paper were made ubham Chaudhuri; the *Review's* referees; seminar participants at the Department of Economics, ell University; and participants at the 1991 Eastern Economics Association meeting.

in improving nutritional levels. Mounting empirical evidence suggests that the income elasticity of household food energy intake in developing countries is lower than had been thought a decade ago (Behrman 1988; Behrman and others 1988; Bouis and Haddad 1988; Alderman forthcoming). There are reasons to suspect that the methodologies used in earlier studies may have overestimated that elasticity (Behrman and Deolalikar 1987; Bouis and Haddad 1988). The downward revision that seems to be warranted is far from negligible. Whereas 10 years ago an income elasticity of energy intake for the poor of about 1.0 would probably not have been seriously questioned, elasticities a good deal less than 0.5 would be considered more in keeping with recent estimates (Behrman and Deolalikar 1987; Bouis and Haddad 1988; Ravallion 1990; Strauss and Thomas 1989; Bhargava 1991).

This substantial downward revision of estimates of how nutrition intake responds to income has some potentially profound implications for development policies. The fight against hunger has been one of the strongest motivations for development, and raising incomes of the poor (through both the growth process and policies to reduce inequalities of income) has long been seen as the main weapon in that fight. The recent evidence has led some observers to suggest that this weapon may be quite blunt or, indeed, virtually useless. Does this aspect of our approach to development policy need a major revision?

This article takes a further look at the question of whether aggregate undernutrition responds to incomes and prices. The point of departure is the realization that nutrient intakes alone are not of interest but, rather, the adequacy of those intakes relative to needs (Ravallion 1990). Although this is obvious at a conceptual level, it is difficult to measure. The methodology proposed here is potentially far more robust to the inevitable errors and unknowns in measuring attainment relative to needs than are previous econometric studies.

Beginning with an econometric model of how nutrient intake distributions vary across regions or sectors of the economy, theoretical results on stochastic dominance are used to infer the effects of changes in incomes and prices on undernutrition. The usefulness of stochastic dominance theory for ranking distributions in terms of some objective function has been known for more than 20 years (Hadar and Russell 1969; Rothschild and Stiglitz 1970), although the relevance to poverty and nutrition analysis has come to be appreciated only quite recently (Atkinson 1987; Foster and Shorrocks 1988; Kakwani 1989). In assessing impacts on undernutrition, the dominance approach, outlined in section I, has the advantages over past methods that it uses all of the information available on the distribution of nutrient intakes, and it places far fewer restrictions on the unknown distribution of individual nutrient requirements. It can also allow inferences that are more robust to intake measurement errors and to arbitrary choices about the specific measure of undernutrition.

This study combines the dominance approach with an econometric model of intake distributions, thus permitting dominance tests of the comparative static effects on undernutrition of changes in the explanatory variables of the intake

model. In section II the methodology is used to explore Indonesia's progress in reducing undernutrition during the mid-1980s, especially the regional dimensions of that progress and the role played by changes in incomes and prices.

In section III other questions concerning the distribution of the benefits of growth are explored: How responsive is undernutrition to changes in income inequality? Will growth that is associated with an increase in inequality be able to reduce undernutrition? If income redistribution entails a "growth tradeoff" (due to, for example, adverse effects on aggregate savings), how severe would that tradeoff have to be to neutralize the effect of greater equity on undernutrition? Section IV summarizes the conclusions.

I. Approaches to Assessing Impacts on Undernutrition

The biggest single problem in past econometric studies of nutrition is that one rarely knows what nutrients the surveyed individuals require to maintain normal physiological functions without symptoms of deficiency. Requirements may vary widely among individuals because of differences in metabolic rates at rest and activity levels. Two sources of variability can be distinguished: interpersonal variability (genotypic variations around the assessed requirements of some reference person) and intertemporal variability for a given person, which has been interpreted as the outcome of physiological regulatory mechanisms influencing energy utilization in the human body (as in, for example, Sukhatme 1978 and Srinivasan 1981).

This latter source of variability has led Sukhatme and others to advocate a cutoff point in measuring undernutrition that is well below stipulated requirements for a given person. Although the existence of intrapersonal variability may lead one to use caloric cutoff points that differ from the stipulated requirements, whether the cutoff point should be set lower or higher will depend on the relative importance attached to "type 1 errors" (incorrectly classifying the person as undernourished) and "type 2 errors" (incorrectly classifying the person as adequately nourished) (Dasgupta and Ray 1990; Kakwani 1989). Advocacy of a lower cutoff point can be interpreted as a judgement that type 1 errors matter more than type 2 errors. This is clearly a questionable judgement (Dandekar 1981; Osmani 1987; Dasgupta and Ray 1990). For this discussion the source of the variability is not specified, so that the distribution of individual-specific requirements at the survey date could reflect either interpersonal or intertemporal variability.

Uncertainty about nutritional requirements is not the only problem. Nutrient intakes, although more readily observed than requirements, are typically measured with error, such as those resulting from imperfect recall of food consumption. In effect this will also mean that the individual requirements specified are incorrect and should ideally be adjusted for the intake measurement error. This measurement problem is another reason for treating requirements as a random variable.

By making explicit assumptions about the interpersonal distribution of nutrient requirements, measures of undernutrition can be readily constructed from household or individual survey data, and, using an appropriate econometric model of intake determination, income and price effects can then be measured (Ravallion 1990). But the assumptions made about requirements are often arbitrary, and they may influence the conclusions drawn from such an exercise. Fortunately, the main interest is often in the qualitative effects of changes in incomes and prices on undernutrition: is some policy combination, interpreted through a set of price and income changes, moving undernutrition in the right direction? In this case there is an alternative approach utilizing stochastic dominance tests, which place far fewer restrictions on the unknown distribution of individual requirements. With this approach each person may have a different requirement, and the distribution need not have any recognizable form.

Although this approach allows a far more general class of possible distributions of requirements, it still imposes two potentially important assumptions about the distribution. First, as is commonly assumed following Sukhatme (1961), intakes and requirements are taken to be independently distributed. One would suspect that the influence of common factors such as age, weight, and activity levels would lead intakes and requirements to be positively correlated. Nonetheless there is some evidence suggesting that there is little or no correlation between intakes and requirements (Sukhatme 1961), and Kakwani (1989) has further found that estimates of the proportion of the population of India deemed to be undernourished are quite insensitive to the assumption one makes about the correlation between intakes and requirements. Second, it is assumed that the distribution of requirements does not change over time or, in the specific cases here, is independent of incomes and prices. Whether that is true will depend on how the changes in incomes and prices occur. If intakes improve as a result of greater effort leading to higher incomes, for example, then undernutrition need not improve. The reasonableness of these assumptions should be kept in mind when interpreting the empirical results below.

Under these assumptions one can invoke well-known theoretical results on stochastic dominance to at least partially order the intake distributions in terms of any well-behaved measures of undernutrition. To illustrate, suppose the intake distribution shifts from state A to state B, as a result of (say) changes in the income distribution; state B, for example, may be reached after economic growth in state A. If intakes increase for all individuals in the move from A to B and there is no change in any individual's requirement, then undernutrition must fall.

But that is an unnecessarily demanding test for a reduction of undernutrition. A more useful test can be constructed as follows. Let $F_t(z)$ denote the proportion of the population that fails to reach a given intake level z in state t. In the vocabulary of poverty measurement, $F_t(z)$ is the headcount index of undernutrition when a single requirement cutoff point is set at z. As z is allowed to vary over its entire range, $F_t(z)$ traces out the cumulative distribution function of

intakes in state t. If $F_A(z)$ is below $F_B(z)$ at all intake levels (or, more precisely, $F_A(z)$ is nowhere above $F_B(z)$ and at least somewhere below), then the proportion of the population that is undernourished will be lower in A than B. This is called a first-order dominance test for comparing undernutrition in two states.

First-order dominance can be an extremely useful test for determining whether there is more poverty in one state than another for any unknown, but fixed, poverty line (Atkinson 1987; Foster and Shorrocks 1988). Furthermore the underlying theoretical result that supports this test can be readily generalized to accommodate any fixed distribution of poverty lines or (as in this application) nutrient requirements (Kakwani 1989). This is the case because our best estimate of the proportion of the population that is undernourished for any distribution of requirements is the expected value of $F_t(z)$, where the expectation is taken over that distribution of requirements. If two intake distributions have the same requirement distribution and first-order dominance holds, then the expected value of $F_A(z)$ must exceed $F_B(z)$ when both expectations are evaluated over the distribution of requirements. Undernutrition is higher in state A than in state B.

First-order dominance of one intake distribution over another also implies an unambiguous ranking of the two distributions in terms of a broader class of undernutrition indicators than the simple headcount index. Let the level of undernutrition of a person with intake x and requirement z be $u(x,z)$, which is positive for $x < z$, but zero otherwise. For the headcount index, $u(x,z) = 1$ for $x < z$, but zero otherwise. Thus the headcount measure is insensitive to differences in the severity of undernutrition. A more general assumption is that the function u is strictly decreasing in intakes and increasing in requirements for all $x < z$ and vanishes at $x = z$. Aggregate undernutrition is then the expected value of $u(x,z)$ over the relevant intake and requirement distributions. For example, if $u(x,z) = (1 - x/z)^\alpha$ for $x < z$ and any positive parameter α, then we have the measure of undernutrition proposed by Kakwani (1989). (This generalizes the Foster, Greer, and Thorbecke 1984 class of poverty measures so as to allow the poverty line to be a random variable.) It can be readily demonstrated that first-order dominance then implies an unambiguous ranking of two distributions in terms of aggregate undernutrition. This result holds for any functional form for $u(x,z)$ with the aforementioned properties and any fixed distribution of requirements.

If the first-order dominance test is inconclusive (with cumulative frequency distributions intersecting at one or more interior points), then some requirement distributions and some reasonable measures of undernutrition will rank the two intake distributions inconsistently compared with others. Here an analogous second-order dominance test may prove useful. In particular suppose one considers only the undernutrition measures that are strictly decreasing in intakes (thus excluding the headcount index). It can be shown that if the area under the cumulative distribution function of intakes in state A is less than that in state B at all intake levels, then aggregate undernutrition is lower in A than B for any

requirement distribution. Kakwani (1989) proves this for the aforementioned class of undernutrition measures based on the Foster-Greer-Thorbecke functional form. Slightly modifying the argument in Atkinson (1987), one can show that the claim also holds for a broad class of other functional forms for $u(x,z)$. If this second-order test fails, one can invoke higher-order dominance tests, requiring the further restriction of the class of admissible measures of undernutrition, or one can restrict the set of admissible requirement distributions.

In applying the dominance approach, the distribution of intakes at each of two dates can simply be compared, as in Ravallion and Huppi (1991). This is useful in assessing overall progress in alleviating poverty or undernutrition. However, any observed change is likely to be the result of changes in many factors, and a simple comparison of distributions reveals nothing about the relative importance of different factors. A more illuminating approach, which is followed here, first models the intake distribution as an explicit function of the variables of interest and then uses this model to conduct dominance tests. (For another, somewhat different, example of this approach, see Ravallion and van de Walle 1991.)

This methodology does not yield estimates of intake elasticities at the household or individual level, such as those widely discussed in the recent literature. This may be viewed as a limitation of the methodology, but, as argued elsewhere, this limitation is not serious, since the individual intake elasticity can be a poor guide to the effects of policy on undernutrition (Ravallion 1990). If quantitative results are desired, a better measure is provided by the elasticities of the cumulative intake distribution $F(x)$ with respect to the variables of interest. These are derived naturally in the proposed methodology.

II. An Application to Indonesia

The data tapes of the National Socioeconomic Surveys (SUSENAS), conducted by Indonesia's Central Bureau of Statistics, estimate household energy intakes by applying caloric unit values to the reported consumption of about 170 food and beverage items during the previous week. Most of the recent literature focuses on caloric intake as the sole determinant of nutritional well-being, and that approach will be followed here. The 1984 and 1987 SUSENAS consumption surveys contain stratified random samples of 50,000 households in each year. Population estimates were derived using the household-specific inverse sampling rates given on the SUSENAS data tapes.

Because respondents were asked about expenditures on foods not prepared at home but not about the quantities consumed of such foods, the survey underestimates energy intakes. Food prepared outside the home is an important source of calories in Indonesia, particularly for certain socioeconomic groups such as those living alone. Although undernutrition will probably be overestimated, it is less clear how this would affect distributional comparisons over time and across regions. Average energy intake from foods prepared away from home is esti-

mated to be 128 calories per person per day in 1987 and 112 calories per person per day in 1984 (Biro Pusat Statistik 1989a). Assuming that calories obtained from this source increase with income, this omission could cause the income response of true intakes to be underestimated. Although *it is reasonable to* assume that "eating out" is a normal good, however, it is also a heterogeneous good, and casual observation suggests that food from street stalls is popular among both poor and not-so-poor in Indonesia. Furthermore, while expenditures on prepared foods increase with income (Biro Pusat Statistik 1989b), it is less likely that intakes from this source will do the same, as the unit values will undoubtedly also increase. The exclusion of calories consumed outside the home is not likely to bias the elasticity estimates below.

The SUSENAS indicates a marked improvement in the distribution of energy intakes in Indonesia during the 1980s. Ravallion and Huppi (1991) construct the cumulative frequency distributions of intakes for 1984 and 1987 from the SUSENAS consumption surveys. Their results indicate that the 1987 intake distribution lies below that for 1984 up to a high intake level. First-order dominance therefore holds over a very wide range of potential cutoff points. Furthermore the second-order dominance condition holds over the whole range of intake cutoff points. A wide range of measures of undernutrition will thus indicate an improvement between 1984 and 1987, irrespective of the distribution of requirements, assuming that the latter did not change. This conclusion is unlikely to be affected by excluding foods prepared away from home.

Do these aggregate results mask significant regional variation? After dividing the 27 provinces of Indonesia into urban and rural areas, which produced 52 regions (Jakarta being solely urban, and East Timor having only rural observations in the sample), the intake distributions were constructed for each region, and the dominance tests applied. An unambiguous improvement was indicated for 29 regions, while an unambiguous worsening was indicated for only three (the rural areas of Riau, Bengkulu, and Maluku). The first-order test was ambiguous for the remaining 20 regions. For each of these the 1987 distribution crossed the 1984 just once from below, that is, there was an improvement at the lower end of the distribution. The average crossover point was at approximately 1,950 calories per person per day (probably around 2,100 with food eaten away from home). The second-order test resolved this ambiguity for 12 regions, all but one of which (urban Central Sulawesi) showed an unambiguous improvement. The regions for which the second-order test was ambiguous were urban and rural Aceh, rural West and South Kalimantan, rural South-East Sulawesi, and the urban areas of Riau, Jambi, and North Sulawesi.

Although first-order or second-order dominance tests thus indicate a reduction in undernutrition for about 80 percent of the regions, the quantitative extent of this improvement varies considerably across regions. Table 1 gives the mean and standard deviation of the percentage change in the cumulative distribution of intakes between 1984 and 1987 for selected intake levels. Both the average percentage decline and its standard deviation tend to be higher at the

Table 1. *Average Percentage Change in the Proportion of the Population below Selected Calorie Intake Levels for 52 Regions of Indonesia, 1984–87*

	Calories per person per day					
	1,200	*1,400*	*1,600*	*1,800*	*2,000*	*2,200*
Mean percentage change	−46.8	−25.5	−12.9	−9.3	−5.2	−2.0
Standard deviation	38.7	31.4	24.7	18.5	14.2	12.6
Coefficient of variation	0.8	1.2	1.9	2.0	2.7	6.4

Note: Calorie intake excludes energy from food not prepared at home.
Source: Author's calculations based on regional data files constructed from the household-level data tapes of Indonesia's National Socioeconomic Surveys for 1984 and 1987.

lower end of the distribution, although the coefficients of variation increase with the intake level. The large standard deviations signal the magnitude of differences in nutritional improvements.

To what extent can differences in income growth explain this regional diversity in the rate of nutritional improvement? Table 2 reports the elasticity of the proportion of the population below various intake levels with respect to income and to total food expenditure (obtained by regressing percentage changes against percentage changes between 1984 and 1987; see Ravallion 1991 for further information). An inverse relationship is evident at all points for both income and expenditure, with the (absolute) elasticity decreasing sharply as the intake level increases. Absolute elasticities are higher for food expenditure than for income at all points. (These elasticities are not comparable with the intake

Table 2. *Income and Expenditure Elasticities of the Proportion of the Population below Selected Energy Intake Levels by Region of Indonesia, 1984–87*

Calories per person per day	*Elasticity with respect to*	
	Income	*Food expenditure*
1,200	−1.190	−1.480
	(0.259)	(0.168)
1,400	−0.633	−0.852
	(0.185)	(0.128)
1,600	−0.435	−0.537
	(0.127)	(0.094)
1,800	−0.355	−0.424
	(0.092)	(0.067)
2,000	−0.228	−0.265
	(0.070)	(0.054)
2,200	−0.155	−0.166
	(0.061)	(0.050)

Note: Elasticities are based on OLS estimation with data as percentage changes over the period. Standard errors are in parentheses; $n=52$. Calorie intake excludes energy from food not prepared at home.
Source: Author's calculations based on regional data files constructed from the household-level data tapes of Indonesia's National Socioeconomic Surveys for 1984 and 1987.

elasticities often quoted in the literature discussed above; they are the elasticities of the proportion of the population below each intake cutoff point rather than the elasticities of intake at that point.)

These results suggest that, first, regions with higher rates of income growth tended also to have greater rates of improvement in the intake distribution. Second, the proportionate shift induced by a given income growth rate tended to be greater at lower intake levels. And third, there is considerable regional variation around these trends. There are, for example, many regions in which the intake distribution noticeably worsened, despite at least modest growth in mean incomes.

This last observation could reflect any number of factors, including higher prices for staple foods, worsening intraregional income inequalities, or deteriorating related health services. To further explore the determinants of the shifts in the intake distributions, the following model was estimated across regions and for each of six values of the intake cutoff:

$$L^i_{87}(x) = Z_i\gamma(x) + \beta(x)L^i_{84}(x) + \epsilon_i(x), \qquad i = 1, \ldots, 52 \tag{1}$$

where

$$L^i_t(x) \equiv \log[F^i_t(x)/(1 - F^i_t(x)] \qquad t = 1984, 1987$$

x is the value of the intake cutoff, and Z denotes a vector comprising mean individual income in the region, mean income squared, the Gini index of inequality in consumption within the region, the average price of rice, and an urban-rural dummy variable equal to 1 if the region is rural. A logit transform is applied to the intake frequencies in equation 1 to avoid truncation in the distribution of the model's dependent variable. The model's random error terms ϵ_i are assumed to include region-specific random effects, entering additively, and the lagged dependent variable is treated as endogenous. The Bhargava and Sargan (1983) maximum likelihood estimation method for dynamic random effects models is appropriate for estimating equation 1 and has been used to obtain the results reported in table 3, below.

The mean income and Gini index were estimated from the household-level data tapes. Income data in the SUSENAS include imputed values of own production. An allowance was made for urban-rural price differences (following Ravallion and Huppi 1991) but not for regional price differences, apart from rice. (There is no satisfactory price index for this purpose for Indonesia. One candidate, a deflator based on the raw data for the consumer price index, was tested as an additional independent variable, but proved insignificant.) Only Gini indexes for total consumption expenditure were included because this probably indicates inequality better than a Gini index based on income. The urban-rural dummy variable may pick up some of the measurement error in the dependent variable associated with the greater importance of energy obtained from prepared foods in urban areas.

Although dynamic effects are allowed in the model, there is little sign of such

Table 3. *Estimation Results from Models of the Determinants of Energy Intake Distributions across Regions of Indonesia, 1987*

	Calories per person per day					
	1,200	*1,400*	*1,600*	*1,800*	*2,000*	*2,200*
Intercept	46.23	20.64	21.65	20.21	−18.68	−39.52
	(0.92)	(2.94)	(0.68)	(0.12)	(2.85)	(4.1)
Rural dummy variable	0.62	0.59	0.89	1.24	1.66	1.96
	(0.27)	(0.38)	(0.12)	(0.14)	(0.56)	(0.35)
Log mean income	−10.07	−4.28	−4.54	−4.29	3.61	8.12
	(0.062)	(0.07)	(0.03)	(0.02)	(0.04)	(0.17)
Log income squared	0.42	0.15	0.156	0.13	−0.28	−0.52
	(0.02)	(0.02)	(0.004)	(0.003)	(0.03)	(0.02)
Log rice price	1.38	0.68	1.03	1.14	1.60	1.59
	(0.14)	(0.22)	(0.11)	(0.03)	(0.17)	(0.86)
Gini index	7.22	6.55	5.12	5.68	7.42	9.29
	(3.51)	(2.40)	(1.33)	(1.23)	(3.47)	(4.32)
Lagged dependent variable	0.31	0.26	0.07	−0.21	−0.52	−0.79
	(0.27)	(0.17)	(0.18)	(0.16)	(0.58)	(0.43)

Note: These are dynamic random effects models, using the Bhargava-Sargan maximum likelihood estimator. Standard errors are in parentheses; $n = 52$. The dependent variable is the logit transformation of the ordinate of the energy intake frequency distribution in 1987. Calorie intake excludes energy from food not prepared at home.

Source: Author's calculations based on regional data files constructed from the household-level data tapes of Indonesia's National Socioeconomic Surveys for 1984 and 1987.

effects in the results, and so these will be ignored in the following discussion. The γ parameters are estimated quite precisely (table 3). Several alternative specifications were also estimated. These (alternately) deleted the lagged dependent variable and the squared income term, treated income as endogenous, and used mean consumption and mean food expenditures as alternatives to income. Similar results were obtained with and without the alternative specifications; details are available from the author. In addition rejection of the null hypothesis that $\beta(x) = 1$ implies rejection of the (nested) alternative specification using proportionate changes as the dependent variable.

The results in table 3 indicate that, for all intake levels considered, the intake distribution is strictly decreasing in mean income (the turning point of the quadratic function of log income is outside the range of the data in all cases) and strictly increasing in both the mean price of rice and the Gini index. Thus first-order dominance is satisfied, implying that undernutrition falls with higher average incomes, lower rice prices, and lower inequality for any (fixed) distribution of requirements.

III. Further Implications of the Empirical Results

Interpretation of the quantitative results is easier if one calculates the implied elasticities of the intake distribution $F(x)$ with respect to each of the variables of interest. Table 4 gives these elasticities of $F(x)$ at the mean of each variable for each value of x. The income elasticity (this time controlling for the other vari-

ables in the above model) is again found to decrease sharply as the intake cutof point rises. The same is true of the elasticities with respect to the Gini index an the price of rice.

Table 5 gives simulated effects on the intake frequency distributions of 1 percent changes in mean income, the Gini index, and price of rice, alternatel all evaluated at variable means. The new intake distribution resulting from an change dZ is estimated by its linear approximation $F_{87}[1 + (1 - F_{87})dZ\gamma]$. Bot income growth and greater equity alleviate undernutrition, and they hav roughly equal proportionate impacts; around the middle of the distribution, 10 percent increase in the mean income is found to have about the same impac on the intake distribution as a 10 percent decrease in the Gini index o inequality.

This suggests a further question: what is the effect on the intake distribution i income growth is accompanied by increased inequality? The correlation coeffi cient between the Gini index and mean incomes across the 52 regions for 1987 i significant and positive at 0.47, although for 1984 the correlation coefficient i just 0.24, and the correlation coefficient between proportionate changes in th two variables between 1984 and 1987 is a weak, although positive, 0.23. Ther is at best a slight hint of a Kuznets relationship in these data. One should b wary of giving this correlation a causal interpretation, but it is still of interest t ask: if income growth occurs with an increase in inequality consistent with th above correlation, would undernutrition still fall? An ordinary least square regression of the Gini index against the log of mean income in 1987 give (t-ratios in parentheses):

$$\text{Gini index} = \underset{(1.679)}{-0.223} + \underset{(3.737)}{0.048}\log(\text{mean income}) \quad (2)$$

with $R^2 = 0.22$. (Using a logit transformation of the Gini index to avoi truncation of the dependent variable yields similar results; see Ravallion 1991.)

Table 4. *Estimated Elasticities of Cumulative Frequencies of Energy Intake at Selected Intake Levels, Indonesia 1987*

Calories per person per day	*Elasticity with respect to*			*Critical income share*[a]
	Mean income	*Gini index*	*Rice price*	
1,200	−1.31	1.86	1.30	0.99
1,400	−0.95	1.44	0.55	0.57
1,600	−0.84	0.90	0.66	0.78
1,800	−0.77	0.74	0.55	0.71
2,000	−0.71	0.66	0.52	0.74
2,200	−0.56	0.55	0.34	0.61

Note: Calorie intakes exclude energy from food not prepared at home.

a. Maximum revenue from rice production as a proportion of total income, which is consistent with positive total elasticity with respect to rice price (that is, allowing mean income to vary).

Source: Author's calculations based on regional data files constructed from the household-level dat tapes of Indonesia's National Socioeconomic Surveys for 1984 and 1987.

Table 5. *Actual and Simulated Cumulative Frequency Distributions of Energy Intakes for Indonesia, 1987*
(percentage of individuals with intakes below each cutoff)

	Cumulative frequency							
		With 10 percent change in						*With 10 percent increase in mean income and an increase of 0.005 in the Gini index*
		Mean income		*Gini index*		*Rice price*		
Calories per person per day	*Actual*	*Increase*	*Decrease*	*Increase*	*Decrease*	*Increase*	*Decrease*	
1,200	6.02	5.23	6.81	7.14	4.90	6.80	5.24	5.43
1,400	19.38	17.53	21.23	22.18	16.58	20.44	18.32	18.02
1,600	35.98	32.94	39.02	39.21	32.75	38.35	33.61	33.51
1,800	52.22	48.21	56.23	56.10	48.34	55.07	49.37	48.90
2,000	67.27	62.49	72.05	71.74	62.80	70.79	63.75	63.27
2,200	78.55	74.14	82.96	82.83	74.27	81.23	75.87	74.89

Note: Calorie intake excludes energy from food not prepared at home.

Source: Author's calculations based on regional data files constructed from the household-level data tapes of Indonesia's National Socioeconomic Surveys for 1984 and 1987.

The implied elasticity at the mean of the Gini index is 0.18; a 10 percent increase in the mean income would only yield about a 2 percent increase in the Gini index at the mean. At this rate of change in inequality, growth would still have a positive effect on undernutrition and almost as strong as the previous partial elasticities had implied. The last column of table 5 gives a revised estimate of the effect of a 10 percent increase in income, incorporating the associated change in inequality. If the growth process induces a Kuznets effect, it would have to be very much more dramatic in its impact on inequality than the above statistical correlation indicates before it would reverse the conclusion that income growth alleviates undernutrition.

It is sometimes also argued that policy interventions aimed at alleviating undernutrition by redistributing incomes in favor of the poor will have an adverse effect on aggregate income, so that the effect on undernutrition would be mitigated. The above results suggest that the percentage decrease in the mean income would need to be of about the same magnitude as the percentage decrease in the Gini index before the redistribution was neutral in its effect on undernutrition. Such a tradeoff may not be plausible.

The empirical results also suggest that undernutrition will be quite responsive to changes in the price of rice, although the elasticity of the intake distribution function falls even more sharply as one moves to higher intake levels (see table 4). Of course this is a partial elasticity, holding both mean income and the Gini index constant. If the share of income obtained from rice production exceeds some critical level, then the income effect of an increase in the price of rice will outweigh the direct adverse effect on undernutrition. (To a first-order approximation, the elasticity of income with respect to the price of rice is simply the share of income from rice production.) The results in table 2 can be used to evaluate this critical income share, assuming that (for lack of any plausible alternative assumption) the price change is distributionally neutral. These income shares are given in the final column of table 4. The critical values are considerably higher than the average income shares from rice production in rural areas of Indonesia. For example, using 1987 data for rural central Java, the average share of income obtained from food grain production (no further disaggregation is possible) is about 20 percent, and this varies little with income. It is evident, then, that an increase in the price of rice will have an adverse effect on both urban and rural undernutrition, allowing for typical income responses.

All these comparative static results hold without qualification for any interpersonal distribution of requirements, provided that this distribution is not also a function of incomes and prices. If requirement distributions shift, then caveats are called for, as noted in the previous section, and the above results may be either strengthened or weakened. Take, for example, an increase in income that arises from an increase in demand for agricultural labor (due to, say, a green revolution allowing multiple cropping). This type of work tends to be strenuous so that the requirement distribution may well shift to the right, reflecting an increase in the mean nutrition requirement. Effects on undernutrition are no

longer obvious. However, income growth associated with labor-intensive industrialization would probably have the opposite effect on the distribution of nutrition requirements; even unskilled modern sector jobs tend to have lower energy requirements than those of the traditional agricultural sector (James and Schofield 1988), thus strengthening the comparative static results on the effects of income growth on undernutrition.

IV. CONCLUSIONS

Although recent household-level consumption studies have revealed much about the micro-level determinants of nutrient intakes, empirical assessments of undernutrition and its response to changes in incomes, prices, or other variables of interest are still confounded by the fact that, while nutrient intakes can be estimated with tolerable precision from such surveys, nutrient requirements are unobserved. Some studies of the impacts of income or other changes on nutritional status have simply ignored requirements, while others have taken a cutoff point that is assumed to be constant for all individuals or constant for all individuals of the same type, with a modest number of types identified. Neither approach is really a convincing foundation for assessing impacts on undernutrition, when (as can be reasonably expected) nutrient requirements for good health can vary widely across individuals and in ways that the researcher can have little hope of knowing. The potential hazards for inference are plain enough.

This article has used an alternative empirical approach that exploits the potential for establishing partial orderings of intake distributions in terms of measures of undernutrition for any unknown but fixed distribution of requirements in a population. Every individual may have a different requirement, which may change over time, although the overall distribution is assumed to be static and independent of the intake distribution. The theoretical results needed to assess comparative static effects on a broad class of measures of undernutrition are then straightforward applications of well-known results from the theory of stochastic dominance.

In implementing the approach here, an attempt has been made to better understand the regional dimensions and economic determinants of the fall in aggregate undernutrition in Indonesia during the 1980s, as found by Ravallion and Huppi (1991). Cumulative distributions of household intakes per person for each of 52 regions of Indonesia have been constructed for 1984 and 1987. Ordinates of these distributions are then modeled as functions of regionally specific variables of interest, focusing particularly on average incomes and rice prices and on the inequality of consumption within regions. Dominance tests have then been applied to assess qualitative effects on undernutrition.

A reduction in undernutrition between 1984 and 1987 is indicated for about 80 percent of the regions by either first-order dominance tests (more than half of the regions) or second-order tests. In modeling the intake distributions across

regions, the empirical results strongly indicate that aggregate undernutrition in Indonesia responds to changes in incomes and prices. These results are consistent with the view that growth in average household incomes under conditions of nonincreasing inequality and reasonably stable prices for food staples has reduced undernutrition in Indonesia. This conclusion is not inconsistent with recent claims that intake responses of individuals to income changes are quantitatively small; the induced shifts on intake distributions can be far from small.

The results also throw light on the relative importance of the identified factors influencing undernutrition. Both growth in average incomes and reductions in inequality will reduce aggregate undernutrition. Even if inequality increases along with income, an implausibly large contemporaneous increase in inequality would probably be required to outweigh the desirable impact of income growth. In addition adverse effects on national income of transfers from rich to the poor would need to be roughly proportional to the associated changes in the Gini index to wipe out the desirable effect on undernutrition of greater equity. Higher rice prices will have an adverse effect on undernutrition, and this will persist under plausible rural income effects associated with the price change. Clearly, the combined nutritional impact of positive (negative) income growth, a decrease (increase) in inequality, and a decrease (increase) in food staple prices can be substantial.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Alderman, Harold. Forthcoming. "New Research on Poverty and Malnutrition: What Are the Implications for Research and Policy." In Michael Lipton and Jacques van der Gaag, eds., *Including the Poor?*

Atkinson, Anthony B. 1987. "On the Measurement of Poverty." *Econometrica* 55 (4): 749–64.

Bhargava, Alok. 1991. "Identification and Panel Data Models with Endogenous Regressors." *Review of Economic Studies* 58 (1): 129–40.

Bhargava, Alok, and J. D. Sargan. 1983. "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods." *Econometrica* 51 (6): 1635–60.

Behrman, Jere R. 1988. "Nutrition and Incomes: Tightly Wedded or Loosely Meshed?" PEW/Cornell Lecture Series on Food and Nutrition Policy. Cornell University, Ithaca, New York. Processed.

Behrman, Jere R., and Anil B. Deolalikar. 1987. "Will Developing Country Nutrition Improve with Income?" *Journal of Political Economy* 95 (3): 492–507.

Behrman, Jere R., Anil B. Deolalikar, and Barbara L. Wolfe. 1988. "Nutrients: Impacts and Determinants." *The World Bank Economic Review* 2: 299–320.

Biro Pusat Statistik. 1989a. *Konsumsi Kalori dan Protein Penduduk Indonesia Per Propinsi 1987*. Jakarta: Biro Pusat Statistik.

———. 1989b. *Pengeluaran Untuk Konsumsi Penduduk Indonesia 1987*. Jakarta: Biro Pusat Statistik.

Bouis, Howarth, and Lawrence J. Haddad. 1988. "Comparing Calorie-Income Elasticities Using Calories Derived from Reported Food Purchases and a Twenty-Four Hour Recall of Food Intakes: An Application Using Philippine Data." Discussion Paper 88. Warwick University, Development Economics Research Centre. Processed.

Dandekar, V. M. 1981. "On Measurement of Poverty." *Economic and Political Weekly* 15: 1241–50.

Dasgupta, Partha, and Debraj Ray. 1990. "Adapting to Undernutrition: The Clinical Evidence and Its Implications." In Jean Dreze and Amartya Sen, eds., *The Political Economy of Hunger*. Oxford: Oxford University Press.

Foster, James, J. Greer, and Erik Thorbecke. 1984. "A Class of Decomposable Poverty Measures." *Econometrica* 52 (3): 761–66.

Foster, James, and A.F. Shorrocks. 1988. "Poverty Orderings." *Econometrica* 56 (1): 173–77.

Hadar, Josef, and William R. Russell. 1969. "Rules for Ordering Uncertain Prospects." *American Economic Review* 59 (1): 25–34.

James, W. P. T., and E. C. Schofield. 1988. "Human Energy Requirements: A Manual for Planners and Nutritionists." Food and Agricultural Organization of the United Nations, Rome. Processed.

Kakwani, Nanak. 1989. "On Measuring Undernutrition." *Oxford Economic Papers* 41 (3): 528–52.

Osmani, S. R. 1987. "Controversies in Nutrition and Their Implications for the Economics of Food." WIDER Working Paper 16. World Institute for Development Economics Research, Helsinki. Processed.

Ravallion, Martin. 1990. "Income Effects on Undernutrition". *Economic Development and Cultural Change* 38 (3): 489–516.

———. 1991. *Does Undernutrition Respond to Incomes and Prices? Dominance Tests for Indonesia*. LSMS Working Paper 82. Washington, D.C.: World Bank.

Ravallion, Martin, and Monika Huppi. 1991. "Measuring Changes in Poverty: A Methodological Case Study of Indonesia during an Adjustment Period." *World Bank Economic Review* 5 (1): 57–82.

Ravallion, Martin, and Dominique van de Walle. 1991. "The Impact on Poverty of Food Pricing Reforms: A Welfare Analysis for Indonesia." *Journal of Policy Modeling* 13 (2): 281–99.

Rothschild, Michael, and Joseph E. Stiglitz. 1970. "Increasing Risk: 1. A Definition." *Journal of Economic Theory* 2 (3): 225–43.

Srinivasan, T. N. 1981. "Malnutrition: Some Measurement and Policy Issues." *Journal of Development Economics* 8 (1): 3–19.

Strauss, John, and Duncan Thomas. 1989. "The Shape of the Expenditure-Calorie Curve." Rand Corporation, Santa Monica, California. Processed.

Sukhatme, P. V. 1961. "The World's Hunger and Future Needs in Food Supplies." *Journal of the Royal Statistical Society* (Series A) 124: 463–525.

———. 1978. "Assessment of Adequacy of Diets at Different Income Levels." *Economic and Political Weekly* 13: 1373–84.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 125-138

Alleviating Transitory Food Crises in Sub-Saharan Africa: International Altruism and Trade

Victor Lavy

The donor community has responded to unexpected or transitory drops in domestic food production in many countries in Sub-Saharan Africa. An empirical framework estimates and analyzes the correlation between this food aid and domestic production. Both emergency food aid and commercial imports are used to offset the effects of negative output shocks in Sub-Saharan African countries—the major recipients of global emergency food aid. On average every one-ton drop in cereal production is offset by the delivery of 0.8 tons of cereal and dairy products from abroad (over four years). Most food aid arrives within a year of a shock, and correlation of shocks over time along with the differences between crop years and calendar years may explain the link between aid and production. The economic and political considerations also determine the global response to the emergency food needs of countries in Africa. The international response is not contingent on the form of government or the level of political and human rights violations. Poorer countries and those with well-established non-emergency food aid programs receive larger amounts of emergency aid when needed.

Food aid has long helped low-income countries to sustain domestic food supplies and cope with transitory food insecurity, and it has also been used to accelerate agricultural development and increase food production. But food aid has been heavily criticized in the past decade on several grounds. First, it has been charged that the international response to food crises is slow, meager, and inefficient. A related accusation suggests that countries granting emergency food aid discriminate on the basis of the political and economic orientation of the recipient country. Second, food aid has been thought to discourage domestic food production, thus leading to long-term dependence on donors. Finally, it has

Victor Lavy is with the Department of Economics at the Hebrew University of Jerusalem. When this article was written, he was a consultant to the Population and Human Resources Department of the World Bank. The article was written under the auspices of the Technical Department of the World Bank's Africa Regional Office as a background paper to the study on Food Aid in Sub-Saharan Africa. The author would like to thank Shlomo Reutlinger for his assistance and useful insight and comments on the paper, the three referees of this journal for their detailed comments and suggestions, and the World Bank Research Committee for its financial support.

been said that, by alleviating shortages, food aid enables countries to postpone or even cancel politically costly economic reforms (Maxwell and Singer 1979; Stevens 1979; Singer and others 1987; Hopkins 1984; Wallerstein 1980; Bhagwati 1986; Srinivasan 1989; Maxwell 1989; Lavy 1990).

This article investigates empirically the validity of some of these criticisms in the granting of emergency food aid to Sub-Saharan Africa during 1979–87. The response of the donor community to unexpected or transitory drops in domestic food production in 26 countries is examined. Food emergencies are characterized here as resulting from unexpected declines in output. However, famines or emergencies are often due to a combination of a breakdown in entitlements and a decline in output rather than just a drop in actual food supply. This study compares the role of food aid and commercial food imports in offsetting these production shocks and covering the shortfall in domestic food supplies. Several hypotheses that postulate the factors determining the response of donors to the needs of different countries are tested.

Section I presents the stylized facts of food insecurity in Africa and documents the relative importance of emergency and total food aid by country. Section II suggests an empirical framework for estimating and analyzing the correlation between food aid and domestic production; the results are presented in section III. Section IV analyzes the economic and political determinants of the global response to the emergency needs of countries in Africa, and section V concludes.

I. Transitory Food Insecurity: The Stylized Facts

Unexpected temporary reductions in food production can have dramatic effects on food consumption. Several mechanisms are available to stabilize consumption, including the commercial importation of food supplies and the granting of emergency food aid. Both options have been used in Sub-Saharan Africa, where domestic food supplies have been unreliable for the past decade.

Table 1 summarizes the coefficients of variation (net of trend) of cereal production during 1970–87. These coefficients are the absolute value of the average annual deviation from the (detrended) mean of production during the sample period, relative to that mean. The countries of the Sahel (Burkina Faso, Ethiopia, Mali, Mauritania, Niger, Somalia, and Sudan) had the largest annual fluctuations, followed by Botswana, Madagascar, and Tanzania. The coefficients reported are averages; the variability in the latter years was much higher than average, because of the 1983–85 drought. Output for even the most stable food producers in Africa fluctuates more sharply than in most other developing countries (World Bank 1988).

Cereal production in Sub-Saharan Africa was not only highly variable but also declined per capita during the sample period. Table 2 gives regional annual growth rates for grain production and population from 1970 to 1982. East and West Africa had the highest rates of population growth and the lowest rates of cereal production growth, which resulted in a decline in per capita cereal pro-

Table 1. *Statistics on Food Production, Aid, and Imports, 1970–87*

Country	*Variability of food production (coefficient of variation)*	*Average emergency food aid as a percentage of total food aid*	*Average total food aid as a percentage of food production*	*Average commercial food imports as a percentage of food production*
Angola	21.3	50	9	84
Burundi*	21.8	9	10	4
Benin*	22.6	18	2	19
Burkina Faso*	39.2	25	3	8
Botswana	77.6	13	195	808
Central African Republic	15.9	10	2	17
Chad	20.3	59	8	10
Cameroon*	13.8	21	1	18
Comoros*	20.0	5	19	124
Congo*	35.8	0	10	862
Cape Verde	99.8	77	1,450	219
Ethiopia	34.6	59	9	8
The Gambia	27.7	34	10	57
Ghana	23.9	18	7	30
Guinea-Bissau	56.3	50	13	24
Côte d'Ivoire*	23.7	20	1	26
Kenya	32.4	23	2	11
Lesotho*	26.4	6	20	82
Liberia*	22.8	0	0	38
Madagascar	26.4	27	1	10
Mauritius	105.7	2	805	9,530
Mauritania	43.4	54	111	525
Mali	36.8	31	6	13
Malawi	33.5	20	1	2
Mozambique	13.7	68	9	60
Rwanda	21.1	23	3	6
Senegal	29.9	45	10	65
Sierra Leone	12.2	5	2	18
Somalia	42.6	81	29	71
Sudan	48.8	60	6	20
Tanzania	52.8	26	3	8
Togo*	18.4	8	2	19
Uganda	33.9	47	1	3
Zaire	25.7	46	1	35
Zambia	25.1	67	1	23
Zimbabwe	43.8	87	1	6

* These countries are not included in the empirical analysis in section II.

Source: Author's calculations based on FAO data.

duction of 2.2 and 0.8 percent a year, respectively. Positive trends were reported in all other regions, with cereal production up 1.8 percent both in East Asia and the Pacific Rim and in the industrial economies.

Africa's negative growth rates during this period were somewhat offset by an increase in food aid and commercial food imports, which augmented total do-

Table 2. *Growth of Cereal Production and Population by Region, 1970–82*
(average annual percentage change)

Region	*Cereal production*	*Population*	*Cereal production per capita*
World	2.3	1.8	0.5
Industrial economies	2.5	0.7	1.8
East Africa	0.8	3.0	−2.2
West Africa	1.9	2.7	−0.8
East Asia and Pacific	3.5	1.7	1.8
South Asia	2.7	2.4	0.3
Latin America	3.2	2.4	0.8

Source: World Bank (1986).

mestic production by an average of 12 and 15 percent in East and West Africa, respectively. As seen in table 1, Botswana, Cape Verde, Mauritania, Mauritius, and Somalia depended heavily on food aid, and Botswana, Cape Verde, Comoros, Congo, Mauritania, and Mauritius depended heavily on commercial food imports. Cross-section correlations indicate that countries with a high coefficient of variation of production receive more aid (as a percentage of total domestic production). The simple correlation coefficient between the variation of production and the ratio of food aid to domestic production is 0.45. This result is even stronger when the ratio of total food aid is replaced by emergency food aid. It appears that low domestic output triggers donor aid—mostly emergency help. Countries that are receiving food aid, and therefore are experiencing food shortages, also tend to import part of the food deficit. The simple correlation coefficient between the share of commercial food imports and the share of total food aid in production is 0.25, and, when emergency food aid is substituted for total food aid, the correlation is 0.31.

II. The Empirical Framework

This section describes the relationship between transitory changes in domestic production and emergency food aid or commercial food imports. Since negative shocks to domestic food production are assumed to be exogenous and are the primary trigger of international emergency food aid, which is assumed to be endogenous, the potential simultaneity between food aid and food production is not problematic. The same applies for transitory output shocks and their effect on commercial food imports. Here, however, negative as well as positive shocks can lead to a change in imports. As a result trade can serve as a symmetric stabilizer of food consumption, whereas emergency aid is used only as a buffer during food shortages. The trade mechanism works mainly through changes in imports from some average level, since food exports are negligible for most African countries. Some exceptions, such as Botswana, Kenya, Malawi, and Zimbabwe, are food exporters on either a continuing or at least sporadic basis. Also some food aid is in fact being given by way of "triangular transaction," for

example, the United Kingdom gives food aid under the International Food Convention through purchases of maize from Zimbabwe that it sends to Zambia.

More formally, let y_t denote the transitory component of food production in year t, E_t denote emergency food aid, and M_t denote emergency food imports. The last is the deviation of commercial food imports from trend and excludes emergency food aid. Response functions for aid and commercial imports in year k have the form

$$E_t = \begin{cases} \sum_{t=0}^{k} \beta_t y_t + u_t, & \text{if } y < \bar{y} \\ 0 & \text{otherwise} \end{cases}$$

$$M_t = \begin{cases} \sum_{t=0}^{k} \tau_t y_t + \epsilon_t, & \text{if } y < \bar{y} \\ \sum_{t=0}^{k} \delta_t y_t + \mu_t, & \text{if } y > \bar{y} \end{cases}$$

and $\bar{y}$ is a threshold below which an output shock will trigger an inflow of food aid or commercial food imports. This threshold is probably different in the two response functions, may not be equal to zero, and may vary from country to country. The emergency aid response model can be treated as a limited dependent variable model, with emergency aid taking only nonnegative values. Emergency food aid in any year may be positive even if the transitory output shock in the same year does not cross the threshold, because the level of emergency food aid in a given period includes a lagged response to previous shocks.

An alternative view of the response model is a switching regime model. One regime is the case of large negative output shocks triggering current (and future) food aid. The other is the case of small negative shocks or positive shocks that do not lead to aid flows. Several difficulties arise in estimating such a model. First, since the threshold level is unknown, the switching model is with unknown regimes. Second, the transitory food production shocks are unobserved and must be estimated. Third, the form of the emergency food aid and emergency food import response functions may be jointly dependent, so that the coefficients should be jointly estimated. Fourth, the lagged response of foreign aid and imports to output shocks may be very important, implying the need to allow for an unrestricted lag structure.

To obtain a measure of the output shocks, an autoregressive model with a time trend for food production is estimated separately for each country. The residuals from these regressions are then used in the two response functions. In the emergency food aid response function, a spline is used so that only the negative shocks are needed to explain the variance of emergency aid, thus constraining the coefficients on positive output shocks to be zero. This means that

the threshold is constrained to zero. This approach is preferred over one that is based on estimating a switching model with an unknown threshold, because it allows greater freedom in exploring both the lag structure and the possibility that the coefficients vary across countries. Some flexibility in the response can be achieved by introducing nonlinearity in the effect of output shocks on emergency food aid. It might be expected that the response for small negative shocks is minimal or even nil, but increases with the severity of the negative shock. Adopting this approach of making the response a quadratic function of the size of the shock allows for different responses for different countries, as long as the countries' shocks are of different size.

The equations for emergency food aid and emergency food import equations can be estimated jointly, using generalized least squares techniques to allow for an unrestricted error structure. As long as the set of explanatory variables is not identical in the two equations, efficiency is gained through joint estimation, which exploits the cross-equation correlation. If the aid response equation includes only the negative output shocks while the import response equation includes both the negative and positive shocks, this condition is met.

The crucial variable needed to estimate the model is emergency food aid. This variable is available for 1979–87 for all the Sub-Saharan countries. Given this short time series, a separate estimation for each country is not possible, and therefore data were pooled. As noted above, differences in the responses across countries may exist. Since the response functions do not include constants, models with individual effects (different intercepts) are inappropriate. An alternative approach is to allow for varying slopes by country, with response being a function of characteristics of the recipient country. The response functions are therefore augmented with an interaction variable between the output shocks and country characteristics as detailed below.

III. International Altruism: The Extent and Speed of the Response

In the aid data for this study (unpublished data supplied by the World Food Programme), food aid is divided into three categories: emergency aid, program aid, and project aid. Each of these categories is subdivided into cereal and noncereal aid. All figures are in tons of grain or grain equivalent. Since data on Africa often suffer from large measurement errors, price and output data must be interpreted with care. To minimize the use of unreliable data, the sample was limited to 1979–87, for which careful attention was given to the quality of the figures. Because the only extended food output and input data available for most of Sub-Saharan Africa are for total cereal production, these are used to estimate random shocks to domestic production and imports. Production and imports are not an adequate estimate of consumption, however, due to the importance of changes in stocks in agrarian societies. The problem, of course, is that the available stock data usually refer only to public stocks, whereas most stocks are actually held by farmers or consumers. Therefore, the stabilization of consump-

Table 3. *Estimation Results of Response Functions for Cereal and Noncereal Emergency Food Aid*

	Cereal		Noncereal	
	With constant term	*Without constant term*	*With constant term*	*Without constant term*
Constant	−783.0	n.a.	532.4	n.a.
	(0.0)		(1.2)	
Positive output shock				
Contemporaneous	0.027	0.026	−0.001	0.001
	(0.6)	(0.6)	(0.1)	(0.1)
Lagged one year	−0.017	−0.017	0.001	0.001
	(0.4)	(0.4)	(0.1)	(0.2)
Lagged two years	−0.126	−0.126	−0.011	−0.011
	(1.9)	(1.9)	(1.4)	(1.4)
Lagged three years	0.048	0.048	0.003	0.004
	(1.5)	(1.5)	(0.7)	(0.9)
Negative output shock				
Contemporaneous	−0.073	−0.073	−0.008	−0.009
	(1.5)	(1.5)	(1.3)	(1.5)
Lagged one year	−0.276	−0.276	−0.026	−0.027
	(5.9)	(6.0)	(4.4)	(4.6)
Lagged two years	−0.037	−0.037	−0.003	−0.003
	(0.8)	(0.8)	(0.5)	(0.6)
Lagged three years	−0.124	−0.124	−0.013	−0.014
	(2.7)	(2.8)	(2.2)	(2.4)
R^2	0.387	0.471	0.236	0.362
Sample size	126	126	126	126
F-statistic	10.875	15.025	5 826	9.924

n.a. Not applicable.
Note: *t*-statistics are in parentheses.
Source: Author's calculations, based on unpublished data from the World Food Programme.

tion from current supplies is modeled here. And because several countries received no emergency food aid or received aid for only a short period, the sample includes only those 26 countries, listed in table 1 without asterisks, that received some emergency food aid for an extended period. Observations that lacked values for emergency aid were discarded, leaving 126 usable observations.

Response functions for cereal and noncereal emergency food aid were first estimated using the contemporaneous and three lagged values of the output shocks, where both negative and positive shocks were included, as well as a constant term (table 3). As suggested earlier, positive shocks to output should have no effect on emergency aid, and indeed the coefficients on the positive shocks were not statistically significant. The evidence suggests therefore that the addition of contemporaneous and lagged positive output shocks, regardless of the lag length, do not contribute to the explanation of variations in current emergency aid flows. The constant term was also insignificant, meaning that there is no autonomous level of emergency aid.

Table 4 presents the generalized least-squares estimates of the emergency food aid response functions to negative output shocks for cereals and noncereals. Experiments with various lag patterns and length show that all emergency aid is

Table 4. Response of Emergency Food Aid to Negative Output Shocks

	Estimated equation					
	With one lag		With two lags		With three lags	
	With constant term	Without constant term	With constant term	Without constant term	With constant term	Without constant term
Cereals						
Constant	15,223.0	n.a.	5,671.8	n.a.	2,107.4	n.a.
	(1.6)		(0.5)		(0.2)	
Negative output shock						
Contemporaneous	−0.031	−0.046	−0.037	−0.043	−0.005	−0.007
	(1.0)	(1.5)	(1.2)	(1.5)	(0.2)	(0.2)
Lagged one year	−0.275	−0.292	−0.253	−0.257	−0.278	−0.280
	(7.1)	(7.8)	(6.6)	(6.9)	(7.1)	(7.3)
Lagged two years	n.a.	n.a.	−0.102	0.109	−0.071	−0.073
			(2.9)	(3.3)	(1.9)	(2.0)
Lagged three years	n.a.	n.a.	n.a.	n.a.	−0.091	−0.092
					(2.2)	(2.3)
R^2	0.315	0.40	0.355	0.44	0.374	0.46
Sample size	126	126	126	126	126	126
F statistic					19.7	27.78
Noncereals						
Constant	2,873.3	n.a.	2,052.0	n.a.	1,656.3	n.a.
	(2.5)		(1.7)		(1.4)	
Negative output shock						
Contemporaneous	−0.006	−0.009	−0.007	0.009	−0.003	−0.005
	(1.7)	(2.5)	(1.9)	(2.5)	(0.8)	(1.1)
Lagged one year	−0.023	0.027	0.029	−0.023	−0.024	−0.026
	(5.0)	(5.8)	(4.6)	(4.9)	(5.0)	(5.4)
Lagged two years	n.a.	n.a.	−0.009	−0.011	−0.005	−0.007
			(2.5)	(2.8)	(1.2)	(1.5)
Lagged three years	n.a.	n.a.	n.a.	n.a.	−0.010	−0.091
					(2.0)	(2.2)
R^2	0.204	0.312	0.228	0.350	0.245	0.367
Sample size	126	126	126	126	126	126
F-statistic	n.a.	n.a.	n.a.	n.a.	11.131	19.283

n.a. Not applicable.
Note: t-statistics are in parentheses.
Source: Author's calculations, based on unpublished data from the World Food Programme.

received within four years of the initial shock, so that the estimates using three lagged values of output shocks are discussed here. For cereals the contemporaneous partial correlation is negative, but not significantly different from zero. The sum of the coefficients on the contemporaneous and lagged shocks is −0.45, suggesting that every one-ton reduction in domestic grain production is compensated by almost half a ton of grain in the form of emergency food aid. Most of this—more than 60 percent—arrives the year after the negative shock. This lag in response may be the result of sluggish reaction by the donor community, but, since the output shocks may be correlated, they could also induce a lagged effect on emergency food aid. For example, if the output shock is caused by a severe drought that forces farmers to temporarily leave their farms, the following year's crop could also be affected.

The timing of the response for emergency noncereal aid is similar to that for cereals: the contemporaneous response is small, with more noncereal aid arriving in the years after the negative shock. The total response of noncereal aid is smaller, however, with only 0.13 of a (grain-equivalent) ton of noncereals being received in aid in the three years following a negative shock.

For some countries for some years there are several missing values of emergency food aid. They are probably zero, but they may be unreported. In order not to throw the country out of the sample, if each missing value is replaced by a zero value, the sample size increases, and the accumulated aid response declines (to 0.35 for cereals and 0.04 for noncereals).

For these estimates of emergency aid response, recall that the threshold that triggers a response has been constrained to equal zero. This implies that every reduction in output beyond that expected according to the long-term trend is assumed to result in emergency aid. In reality, however, emergency aid is provided only when the reduction in output falls beyond a threshold that is less than zero. If this threshold was allowed empirically to be different from zero, the response would be larger than has been estimated here. Indeed higher responses were estimated when negative threshold levels were chosen, but the choice of any threshold is arbitrary. An alternative to using an arbitrary negative threshold is to allow for nonlinearity in the effect of the negative shocks. When the squared value of the output shock was included as a regressor, its coefficient was negative and significantly different from zero, which implies that the aid response is more vigorous as the negative shocks are more severe, thus leading to a larger total response.

With respect to the sluggishness of the aid mechanism, the crop year in which output is recorded and the calendar year in which aid is recorded do not correspond exactly; some of the aid recorded as arriving in the year following the shock actually arrived contemporaneously. So emergency food aid insulates domestic food grain consumption from random shocks to domestic production to a larger degree than the results here imply. But aid that arrives in the years following a shock may cause problems. If the output shock has a short life, the late emergency aid may actually arrive during a period of normal food production and may depress food prices. This effect on prices will be mitigated, however, if the late emergency aid serves to replenish stocks. The estimates here suggest that this "late" (more than a year after the shock) flow is small relative to both total aid and total production, and thus cannot be very harmful.

Random shocks to domestic production may also induce a stabilizing response from commercial imports as domestic food prices increase relative to food prices elsewhere. This mechanism reduces food shortages in a free trade regime. Because many African governments do not, in fact, allow free trade and control imports and food prices, changes in relative prices may be very limited. The import response may nonetheless occur if governments are willing and able to buy on the world market. Table 5 summarizes the parameters of the import response to production shocks. The dependent variable is the deviation—

Table 5. Response of Cereal Imports to Current and Lagged Production Shocks

	All imports		*Commercial imports*	
	With constant term	*Without constant term*	*With constant term*	*Without constant term*
Constant	−19,705.3	n.a.	−22,300.4	n.a.
	(2.5)		(2.2)	
Positive output shock				
Contemporaneous	0.001	−0.014	−0.036	−0.052
	(0.0)	(0.4)	(0.8)	(1.2)
Lagged one year	−0.043	−0.051	−0.011	−0.022
	(1.1)	(1.2)	(0.2)	(0.4)
Lagged two years	−0.080	0.078	−0.095	−0.095
	(1.5)	(1.5)	(1.4)	(1.4)
Lagged three years	0.050	0.043	0.027	0.019
	(1.9)	(1.6)	(0.9)	(0.6)
Negative output shock				
Contemporaneous	−0.106	−0.091	−0.048	−0.034
	(2.7)	(2.3)	(0.8)	(0.6)
Lagged one year	−0.170	−0.158	−0.288	−0.274
	(5.0)	(4.6)	(6.0)	(5.6)
Lagged two years	−0.054	0.047	−0.063	−0.057
	(1.6)	(1.3)	(1.3)	(1.2)
Lagged three years	0.014	0.026	−0.045	−0.032
	(0.4)	(0.8)	(1.0)	(0.7)
R^2	0.352	0.330	0.424	0.408
Sample size	144	144	126	126

n.a. Not applicable.
Note: *t*-statistics are in parentheses.
Source: Author's calculations, based on unpublished data from the World Food Programme.

positive or negative—from the trend of imports derived from estimating the autoregression of imports for each country. Since the dependent variable is a deviation from trend, one would expect the constant term of the regression to be zero, but, for estimating both commercial imports and all imports, the estimated constant term was significantly different from zero.

Results with and without the constant are reported in table 5. The evidence that trade serves as a symmetric buffer indicates that the trade buffer is stronger when there are shortages (through imports) than when there are surpluses. Coefficients on positive shocks are generally negative, as expected, but are not statistically significant. In the case of negative production shocks, shocks lagged more than one period do not have significant coefficients, and commercial imports do not respond contemporaneously to a negative shock. The response of commercial imports in the period following the negative shock is, however, strong and significant. Again this apparent lag may result from the difference between the calendar year and the crop year.

IV. Economic and Political Determinants of Donor Generosity

The global response to the need for emergency aid varies, leading to the hypothesis that the total flow of emergency aid from all donor countries and

international organizations may be influenced by political and economic considerations. To identify the factors that determine the global response to food needs in Sub-Saharan Africa, the aid response equation was reestimated with the addition of interaction terms between current negative output shocks and other political and economic variables that may reasonably be thought to affect the responsiveness of donor countries.

Several hypotheses are tested here. The first is that poorer countries receive more generous responses. The poverty level of a country is measured by per capita income and food consumption. The second hypothesis is that donors may be more responsive to the needs of countries that already receive large amounts of food aid. This hypothesis is based on the assumption that established food aid flows have organized channels of transmission, making emergency aid more effective and less costly. A related factor is the cost of transportation, which in Africa is mostly a function of the land component of transportation. Inland countries may receive less aid because transportation costs are higher for them.

The third hypothesis is that the donors' response may depend on the political orientation of the recipients. The socialist countries in the sample include Angola, Ethiopia, Guinea, Mozambique, and Tanzania, and, in testing this hypothesis, this group was enlarged to include the mixed socialist countries of Madagascar, Mali, Rwanda, Somalia, Sudan, and Zambia. The fourth hypothesis is that the political rights and civil and economic liberties in the recipient country may influence the donor community. Emergency aid may also depend on the form of government. To test these hypotheses, we used Gastil's (1984) ranking of African countries according to political rights, civil liberties, and state of freedom. The first two indexes rank countries from one (best) to seven (worst), while the third index characterizes countries as not free, partially free, or fully free. Most countries in Africa fall between five and seven on the first two indexes. The exceptions appearing in the data set used here are Botswana, Ghana, Mauritius, Senegal, and Zimbabwe. The countries ruled by a military regime are Central African Republic, Chad, Ethiopia, Ghana, and Mauritania.

Given the extensive multicollinearity between all these variables, their impact on aid response was estimated one at a time. When the effect on the response was suspected to originate from correlation between the interaction term and a third variable, that variable was added as a control in the regression. Table 6 summarizes the qualitative results of including the interaction terms in the aid equation for cereals and noncereals. It indicates the sign of the effect on aid and whether it is different from zero at the 5 percent significance level. The signs of the coefficients are the same for the two types of food aid, which is not surprising given the high correlation between the two. Almost all of the coefficients are statistically different from zero.

For economic determinants the hypothesis that the response is better for poorer countries is accepted for both cereal and noncereal aid. Replacing the gross national product per capita as a measure of poverty with other poverty measures, such as per capita food production or per capita daily cereal con-

Table 6. *Factors Affecting the Response of Aid and Trade*

Factor	Emergency		Commercial food imports
	Cereal	Noncereal	(negative shocks)
GNP per capita	−*	−*	−*
Nonemergency food aid	+*	+*	−*
Access to sea	+*	+*	+*
Socialist	+*	+*	+*
Semisocialist	+	+	+
Military government	+	+*	+*
Political rights	+*	+*	+*
Civil liberties	+*	+*	+*
Status of freedom	+	+*	+*

Note: + means a positive correlation; − is negative. * denotes cases in which the interaction term was significantly different from zero (at $\alpha = 0.05$).

sumption, did not change this result in any significant way. The aid response varied greatly with the average level of nonemergency food aid received (as a share of domestic production); countries with a high share of nonemergency food aid received more aid in crisis situations than did countries with low shares of nonemergency aid. If low-income countries tend to receive more nonemergency food aid, however, the correlation between emergency aid and income per capita could be erroneous given the high collinearity between the two components of food aid.

Nevertheless the nonemergency aid effect in the response equation remained unchanged, even when income per capita was controlled for in the estimation. This robust result can be interpreted as follows: the response to emergency needs is positively correlated with the ease (low cost) of delivering food aid, and the cost of transportation and delivery tends to be lower for countries that receive high levels of nonemergency aid. This conclusion is further supported by the significant positive correlation between good geographical location (access to the sea) and the response to output shocks. Although these results suggest that the emergency needs of inland countries are not met with the same generosity as is extended to those with seaports, it is possible that the emergency needs of inland countries are met by means other than food aid.

The results for the political variables are quite surprising. Countries classified as socialist, with military governments, and with a low score in terms of protecting political, economic, and civil liberties tend to receive more aid in emergency or crisis situations. This tendency is probably a result of the high correlation between some of the other determinants of the aid response and the political and human rights factors. Indeed socialist countries and countries with restricted political and economic freedoms are at the lower end of the income distribution in Africa. They receive more aid because they are poorer, and the international response during crises is not contingent on the form of government or the level of political and human rights violations. Regular flows of food aid, however, are highly negatively correlated with socialist or military governments and with poor protection of civil, economic, and political freedoms. These results, how-

ever, likely suffer from sample selectivity biases because some countries with severe food crises did not receive any emergency aid and therefore are not included in the sample (for example, Burkina Faso and Congo).

Table 6 shows the results from reestimating the import response function using only the negative output shocks (the positive shocks were not significant) and their interaction with all the political variables. Surprisingly, this estimation is almost identical to the aid equation. Although the market-oriented economies were expected to be more likely to import food for emergency needs, the results here suggest that emergency imports are higher for the socialist countries. This somewhat unexpected result could reflect the fact that the very poor socialist countries finance emergency commercial imports with general financial aid or with balance of payments relief aid.

V. Conclusions

This article has provided empirical evidence that throws doubt on the validity of some often heard criticisms of food aid. Both emergency food aid and commercial imports are used to offset the effects of negative output shocks in Sub-Saharan Africa. Within one year of an unexpected one-ton drop in cereal production, about 60 percent of this shortfall is replaced by increased food aid and commercial imports. Food aid continues to flow in response to the shock for another two years, although at a much reduced rate.

Several studies have interpreted the negative simple correlation between food aid and domestic food production as an indication that food aid depresses domestic production. The results here suggest that this correlation may be due to the contemporaneous response of emergency food aid flows to output shocks. Most food aid arrives within a year of a shock, and correlation of shocks over time along with the differences between crop years and calendar years may explain the link between aid and production without resorting to the discouragement hypotheses. The dynamics of domestic food production cause and trigger the flow of emergency food aid, so these flows should be treated as endogenous. Emergency aid should be netted out of total food aid in an exercise that evaluates the efficiency of food aid.

This analysis of African countries—the major recipients of global emergency food aid—suggests that on average a shock to agricultural output in the form of a one-ton drop in cereal production would lead to a flow of 0.64 tons within a year. Extending the period to four years increases the compensation to 80 percent. In other words, every one-ton drop in cereal production is offset by the delivery of 0.8 tons of cereal and dairy products from abroad. There is, however, a lag in this response over four years, although most of the aid is received in one to two years. This estimate is biased downward, however, since the threshold that was assumed to trigger emergency flows was only zero.

Contrary to criticism, the pattern of emergency aid flows in Sub-Saharan Africa does not provide evidence of discrimination by donors on the basis of

political factors. The international response to food crises in Sub-Saharan Africa is not contingent on the form of government or the level of political and human rights violations. On the contrary, countries classified as socialist with military governments and with a low score in terms of protecting political, economic, and civil liberties tend to receive more aid in emergencies or crises. Poorer countries and those with well-established nonemergency food aid programs receive larger amounts of emergency aid when needed.

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Bhagwati, Jagdish. 1986. "Food Aid, Agricultural Production, and Welfare." In S. Guhen and M. R. Shroff, eds., *Essays on Economic Progress and Welfare*, in honor of I. G. Patel. New Delhi: Oxford University Press.

Gastil, Raymond D. 1984. *Freedom in the World: Political Rights and Civil Liberties, 1983–84*. Westport: Greenwood Press.

Hopkins, Raymond. 1984. "The Evolution of Food Aid." *Food Policy* 9 (4): 345–63.

Lavy, Victor. 1990. "Does Food Aid Depress Food Production: The Disincentive Dilemma in the African Context." PRE Working Paper 460, World Bank, Population and Human Resources Department, Washington, D.C. Processed.

Maxwell, Simon. 1989. "The Disincentives of Food Aid: A Pragmatic Approach." Institute of Development Studies, University of Sussex. Processed.

Maxwell, Simon, and Hans W. Singer. 1979. "Food Aid to Developing Countries: A Survey," *World Development* 7 (3): 225–47.

Singer, Hans, John Wood, and Tony Jennings. 1987. *Food Aid: The Challenge and the Opportunity*. Oxford: Clarendon Press.

Srinivasan, T. N. 1989. "Food Aid: A Cause of Development Failure or an Instrument for Success?" *World Bank Economic Review* 3 (1): 39–65.

Stevens, Christopher. 1979. *Food Aid and the Developing World: Four African Case Studies*. New York: St. Martin's Press.

Wallerstein, Mitchel B. 1980. *Food for War, Food for Peace*. Cambridge: MIT Press.

World Bank. 1986. *Poverty and Hunger: Issues and Options for Food Security in Developing Countries*. Washington, D.C.

———. 1988. *The Challenge of Hunger in Africa: A Call to Action*. Washington, D.C.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 139-154

The Working Behavior of Young People in Rural Côte d'Ivoire

Rob Alessie, Paul Baker, Richard Blundell,
Christopher Heady, and Costas Meghir

One of the major features of structural adjustment is an attempt to reallocate labor—and hence output—through changes in relative prices. This article assesses how price changes affect the working patterns of young people in rural Côte d'Ivoire. The analysis is based on a model of the labor supply of rural households and on the construction of composite price indexes. The data come from the Côte d'Ivoire Living Standards Survey for 1985 and 1986. The panel aspect of the data allows the work choice made in one year to depend on the individual's choice in the previous year. Results indicate that the price of agricultural output generally is a positive incentive on the decision to participate in the labor force. However, this result depends heavily on the employment and education of the individual in the previous period. Those not already working are less likely to respond to favorable movements in the prices of cash crops by entering the work force.

This article reports on research on the determinants of working patterns among young people in rural Côte d'Ivoire. Because young people's decisions about working are closely related to decisions about schooling, the article also focuses on the relationship between work and school. The motivation for this study arises from an interest in the effects of structural adjustment programs on work incentives. Structural adjustment is expected to increase output by altering the incentives facing producers. The effect of improved incentives could, in principle, be measured by directly observing changes in output. The time lag between the change in incentives and the change in output can be substantial, however, especially when tree crops are involved, so that data over a long period may be

Rob Alessie is with the Department of Economics, University of Tilburg, Netherlands; Paul Baker is with the Institute for Fiscal Studies, London; Richard Blundell and Costas Meghir are with the Department of Economics, University College London and the Institute for Fiscal Studies; and Christopher Heady is with the School of Social Sciences, University of Bath. This article is the final report of a project financed by the World Bank on labor market transitions and structural adjustment in Côte d'Ivoire. The work was undertaken while Rob Alessie was visiting the University College London from Tilburg. The authors are grateful to the World Bank for financial support on this project and to François Bourguignon, Paul Glewwe, Valerie Kozel, John Newman, Jacques van der Gaag, and three anonymous referees for their helpful comments.

necessary to implement this approach. Changes in incentives will also affect input use, and even for tree crops these changes will be more immediate and therefore more readily observable in data from a short period. This article concentrates on labor input.

A microeconomic approach that examines the labor supply decisions of individuals and households has several advantages over an aggregative approach that simply focuses on overall labor participation rates. First, the microeconomic approach provides more units of observation and therefore permits the separate identification of the effects of a greater number of determinants of labor supply. Second, this approach utilizes heterogeneity in the population rather than aggregating across groups, so that empirical results are richer. Third, aggregate labor participation rates may mask the dynamics of labor supply. Small net changes in participation between two years may result from large numbers of people joining and leaving the labor force. Understanding these gross flows into and out of the labor force is essential to predicting the effects of policy changes on the overall participation rate.

This article concentrates on young people for two reasons. First, the young undergo more transitions in labor market status than do older cohorts, which suggests that there is a considerable amount of choice about whether or not young people work. This is not to say that the individual exercises the choice. Decisions about work and school may be made jointly with other members of the household or even just by the head of the household. Second, it is possible that the decision to work may be accompanied by a decision to leave school or at least to reduce time in school. In this case the potential long-term cost of reduced education must be considered in addition to the short-run gain of increased output. This long-term cost is likely to be more serious if it is difficult to reenter school. This possibility should be explored in future research.

The factors influencing labor force participation are expected to be different for urban and rural young people, so it is appropriate to examine each sector separately. The rural sector was chosen for this analysis because it is larger and is the sector in which structural adjustment is expected to have the greatest impact on labor supply and output.

The data used in this study come from the Côte d'Ivoire Living Standards Survey (CILSS) for 1985 and 1986. Côte d'Ivoire is particularly suited for this analysis because it has been carrying out a World Bank–sponsored structural adjustment program, which started in 1981 (Deaton and Benjamin 1987; Glewwe and de Tray 1988). For the CILSS half of the households surveyed in 1985 were reinterviewed in 1986, thus providing an opportunity to observe how individuals have adjusted their labor supply in response to changes in the economic environment. These observations of individual transitions permit an analysis of dynamic behavior that cannot be identified in cross-section data and enable an assessment of the influence of past choices on current behavior. This study is among the first to make use of the panel aspect of the CILSS. It therefore

has the added interest of testing whether these data are sufficiently rich and accurate to allow the estimation of behavioral responses to economic change.

Section I outlines the theoretical basis for the labor supply models that are estimated. Section II describes the data, including a detailed explanation of how price indexes for individual households are constructed. Section III presents the estimation results, and section IV summarizes the findings and indicates directions for further research.

I. The Theoretical Model of the Labor Supply of Rural Households

Most of the rural households sampled in the CILSS are engaged in agricultural production, and their labor supply consists almost entirely of work on their own farm. It is therefore appropriate to analyze labor supply decisions in the context of agricultural household models as described in Singh, Squire, and Strauss (1986). These models represent behavior as the result of collective household decisions about production, labor supply, and consumption.

The Separable Model

In general, decisions about household production and labor supply are interdependent and must be modeled simultaneously. However, under the assumption of complete competitive factor and product markets, the decisions have a recursive structure. Production decisions are made on the basis of maximizing the profit of the farm unit, which produces income for the household. Household labor supply and consumption are then chosen to maximize collective household utility, given consumer prices, wages, and other income. This model of labor supply can be expressed as

$$L_s = F(cp, w, I, z) \tag{1}$$

where L_s is a measure of labor supply, cp is an index of consumer prices, w is the market wage rate, I is profit and other nonlabor income, and z is a vector of household characteristics.

In this "separable" model, agricultural producer prices will affect household labor supply only indirectly through nonlabor income. Increased output prices will raise profits and, assuming that leisure is a normal good, may reduce household labor supply. Because of the recursive structure of the model, the household may simultaneously decide to increase output and total labor input in response to increased prices: the difference between total labor input and household labor supply is covered by labor market transactions.

Consider the incentive effects of structural adjustment in the separable model. The suggestion that increased agricultural prices will reduce household labor supply relies on the implicit assumption that wages are constant. In fact one would expect the increased demand for labor that follows from increased agricultural prices to push wages up. These increased wages will have both income

and substitution effects. The substitution effect will tend to increase household labor supply as leisure becomes more expensive. The direction of the income effect will depend on whether the household is a net buyer or net seller of labor. In either case the income effect will be proportional to the extent of net labor purchases (or sales) and so will be small for the typical Ivorian farm household, which buys or sells relatively small quantities of labor. Thus, in contrast to the situation of nonfarm households, the substitution effect of wage increases can be expected to dominate the income effect and provide the main route through which increased agricultural prices might stimulate labor supply.

Because the wage rate is an exogenous variable in equation 1, the effect of output price changes on labor supply cannot be captured by simply estimating that relationship. The analysis requires the addition of a relationship between product prices and the wage rate:

$$w = G(pp, x) \tag{2}$$

where pp is an index of product prices, and x is a vector of other factors affecting wages. The producer price index in equation 2 will be composed of the prices of all crops produced using labor from the local market in order to capture the role of prices in determining the demand for labor. Since many crops have long periods of production, however, labor demand will depend in part on expected future prices for the product. Ideally, therefore, the price index should also incorporate these expectations.

Equations 1 and 2 could be estimated separately and then combined to provide an estimate of the effect of product prices on labor supply. Alternatively the equations could be combined in a reduced form that could be estimated directly:

$$L_s = H(cp, pp, I, z, x). \tag{3}$$

Although the "structural" estimation of equations 1 and 2 would reveal more detailed information about household behavior and labor market response, the estimation of equation 3 will provide the direct answer to the main question concerning the response of labor supply to structural adjustment. Also equation 3 can be estimated without data on market wage rates, which is a considerable advantage because there is little good quality data on local rural wage rates in Côte d'Ivoire. Finally, as will be shown below, the distinction between equations 1 and 3 becomes blurred as soon as the assumption of perfect product and labor markets is relaxed.

More General Models

There are two main ways in which the labor market might violate the assumption of completeness. First, the labor market may be very thin—or even nonexistent—so that households may have difficulty in either buying or selling labor. The CILSS data suggest that labor market activity in Côte d'Ivoire is rather limited. Landless laborers frequently obtain access to land through sharecrop-

ping rather than through wage labor. Second, hired labor may not be a perfect substitute for household labor because of the need for greater supervision. It is very difficult to obtain empirical evidence about the substitutability between household and hired labor, but the possibility of their being less than perfect substitutes has theoretical plausibility.

Both of these violations imply that household labor supply is not a function of the market wage, but of a shadow wage reflecting the marginal value of extra household labor. This shadow wage will be influenced by the market wage, if there is a labor market, but will depend also on factors that influence the supply and demand for labor within the household. On the demand side, output prices and the amount of available land would be positively related to the shadow wage. On the supply side, the number of household members who are prepared to work would be negatively related to the shadow wage. The number available for work is endogenous, however, so that the nonseparable equivalent to equation 1 is

$$L_s = f(cp, I, w, pp, \text{LAND}, z) \tag{4}$$

where LAND is the quantity of land available to the household.

Two of the variables in equation 4 have a somewhat different interpretation here. First, farm profit is no longer well defined because of the difficulty of valuing household labor. It is possible to define a shadow profit, based on the shadow wage, and its value will depend on the variables already included in equation 4. There is therefore no need to include a separate term for profits, so that I simply denotes nonfarm income. Second, the producer price index pp should now represent the current and expected future prices of goods produced by this particular household, because those prices determine the demand for labor by the household. This suggests that each household should have its own producer price index, with weights reflecting the relative importance of the different crops in its output mix. Once again, if there is a labor market, it is possible to substitute for w, using equation 2 to obtain the reduced form:

$$L_s = h(cp, I, pp, \text{LAND}, z, x). \tag{5}$$

Strictly speaking there should be two producer price indexes, one specific to the household and the other representing labor demand in the whole local labor market. In the empirical work the household-specific index is as described in section II, whereas differences in local labor markets will be captured by regional dummies. Also it is possible that the area of land farmed is also a choice variable of the household. In that case a full analysis of labor force participation would also require modeling land area decisions.

The introduction of producer prices into equation 4 is the most interesting consequence of using the nonseparable model. The absence of producer prices from equation 1 constituted the major difference between it and the reduced-form equation 3, which reflects the fact that producer prices have no direct effect

on labor supply in the separable model. In the more general model, producer prices have a direct effect on household labor supply, in addition to any effect through the market wage rate.

Modeling Individual Labor Supply

The models presented so far describe the determinants of the household decision about overall labor supply. This article, however, concentrates on the labor supply of only part of the household: young people. It is therefore necessary to modify the earlier analysis to focus on the labor supply of a subset of the family. The main consideration is the effect of interdependent behavior of the household members. In particular the working status of other family members will affect the labor force participation of young people for three reasons. First, the working status of other family members will affect the total income of the household. Second, for a given household size, the number of members at work will affect the need for young people to perform household duties and so stay out of the labor force. Third, in the nonseparable model, the number of household members working on the farm will affect the marginal productivity of additional workers.

When all household members have the same skill, larger numbers of household workers would be expected to reduce the marginal product of the young person. If the young people are less skilled and require supervision, however, the availability of supervisors might increase their marginal product. These complications require the addition of the labor force participation of other household members to equation 5. This is clearly an endogenous variable, and the estimation procedures below account for this.

II. The Data and the Household-Specific Output Price Indexes

The data used for this study are from the Côte d'Ivoire Living Standards Survey (CILSS), for which two years of data (1985 and 1986) are currently available. The survey, conducted annually, covers some 1,600 African households a year in communities selected to be nationally representative. The survey design uses a two-stage sampling procedure, with 100 primary sampling units selected from a list of cities, villages, and rural areas sorted by region and type of location. Within each unit 16 households are interviewed. In half of the units all households are surveyed again the following year, while for the remaining half a new sample of households is drawn. In the following year these new households are reinterviewed, and the existing panel households are discarded. The survey thus provides a rolling panel of 800 households. In practice some of these households will move between the sampling years or may refuse to participate in the second interview, thus reducing the panel size. In a preliminary matching of households for 1985 and 1986, 715 households were reinterviewed, providing information on 1,930 individuals.

The CILSS consists of three questionnaires: a household questionnaire, a com-

munity questionnaire, and a price questionnaire. The household questionnaire, which forms the main part of the survey, requests detailed information on household income, expenditures, education, health, labor force participation, and other household and individual characteristics. With regard to changes in individuals' work activity, the survey asks for current employment status, kind of employment (self-employed or a wage laborer), occupation and industry, and hours and weeks at work. Agricultural households are asked for details about land holdings, type of agriculture, crops grown, and livestock raised. The survey also distinguishes between production for own consumption and for market, so that the value of home production can be estimated.

The village questionnaire is completed in all rural clusters to collect information on community characteristics common to all households within a cluster. In particular the questionnaire requests information on local economic activities, economic infrastructure, and social services. The price questionnaire, which is completed in all clusters, collects information on a limited number of basic food items and a few nonfood items.

A Picture of Work Patterns and Transitions

Before considering the labor market activity of the chosen subsample, the general pattern of labor market activity of all panel individuals living in rural areas should be considered. Table 1 shows, for both years and by age group, the proportion of individuals within the panel who reported having worked in the preceding 12 months.[1] Roughly 95 percent of individuals between the ages of 24 and 60 report some labor market activity. There is an increase in overall reported labor market activity in 1986, which is concentrated in age groups with the lowest percentage of workers, namely individuals under 12 or over 60. A somewhat curious feature is the slight decrease in work activity for the 12-to-18 age group.

Table 1. *Labor Force Participation of Rural Individuals, by Age Group, 1985 and 1986*

Age group (in 1985)	*Percentage who worked in*		*Number of observations*
	1985	*1986*	
7–11	22.5	41.3	417
12–18	71.9	69.8	391
19–23	86.2	90.0	130
24–30	94.2	96.7	154
31–40	96.8	96.7	217
41–50	97.4	96.1	230
51–60	95.1	94.5	182
Over 60	66.5	73.2	209
Total	71.4	76.0	1,930

Source: World Bank data.

1. Many panel individuals reported age changes between interviews of more than one year. In selecting the sample, only those individuals for which the absolute value of the difference in age from one interview to the next is less than four years are retained. Some 200 individuals were discarded due to this selection rule.

Table 2. *Work Status Transitions, by Age Group, 1985 and 1986*

Age group (in 1985)	*Number of people who worked*				
	Neither year	*1985 only*	*1986 only*	*Both years*	*Total*
7–11	202	43	121	51	417
	(48.4)	(10.3)	(29.0)	(12.2)	(100.0)
12–18	53	65	57	216	391
	(13.6)	(16.6)	(14.6)	(55.2)	(100.0)
19–23	9	4	9	108	130
	(6.9)	(3.1)	(6.9)	(83.1)	(100.0)
24–30	3	2	6	143	154
	(1.9)	(1.3)	(3.9)	(92.9)	(100.0)
31–40	4	4	3	206	217
	(1.8)	(1.8)	(1.4)	(94.9)	(100.0)
41–50	3	6	3	218	230
	(1.3)	(2.6)	(1.3)	(94.8)	(100.0)
51–60	5	5	4	168	182
	(2.7)	(2.7)	(2.2)	(92.3)	(100.0)
Over 60	45	11	25	128	209
	(21.5)	(5.3)	(12.0)	(61.2)	(100.0)
Total	324	140	228	1,238	1,930
	(16.8)	(7.3)	(11.8)	(64.1)	(100.0)

Note: Percentages of age groups are in parentheses.
Source: World Bank data.

To examine movements of individuals between the states of working and not working, table 2 shows the number and proportion of individuals according to their working status in each year. Between the ages of 19 and 60 the number of individuals leaving or entering the work force is very small. The vast majority of changes in working status are in fact concentrated among individuals aged 18 and under in 1985. The subsample to be used in the empirical work below is comprised of 391 panel individuals aged 12 to 18 in 1985 and living in rural areas. Of this group 122 individuals, or 31.2 percent, report movements between working states across the two years of data, although labor force participation changes by only 2.1 percentage points. The small net change clearly masks large gross flows.

Young people's decisions on work activity may be closely related to decisions on schooling. The school system in Côte d'Ivoire is modeled on the French system, inherited under colonial rule. Six years of elementary education lead to the Certificat d'Etudes Primaires (CEPE), which is awarded on the basis of a nationwide examination. The CEPE is a prerequisite for entrance to secondary school, but, since secondary school places are scarce, the score required to gain placement often exceeds that required to obtain the CEPE certificate. Four years of lower secondary education lead to the Brevet d'Etudes du Premier Cycle, which, if successfully completed, allows the student to enter three years of upper secondary education leading to the Baccalaureate. Alternatively, students with a CEPE certificate or who successfully complete some or all of their lower secondary education can enter various training programs. In the Ivorian education

Table 3. *Work Status and School Attendance, 1985 and 1986*

Age (in 1985)	Percentage attending school in 1985	Percentage attending school in 1986	Percentage working in 1985	Percentage working in 1986	Number of observations
12	68.4	64.5	46.1	48.7	76
13	57.7	53.9	64.1	59.0	78
14	40.0	36.5	73.1	59.6	52
15	29.1	25.5	87.2	83.6	55
16	29.3	21.9	82.9	80.5	41
17	12.8	10.3	84.6	92.5	39
18	12.0	10.0	86.0	88.0	50
Total	40.2	36.3	71.9	69.8	391

Source: World Bank data.

system many students at all levels repeat grades, thus the number of years for which students are enrolled will on average exceed the minimum numbers outlined here.

Table 3 shows the percentage of individuals who attend school and/or work at any time in the 12 months before the survey, according to their age in 1985. Work activity reported is concentrated among those individuals aged 15 to 18 in 1985. Overall, school attendance falls steadily between the ages of 12 and 16, with a distinct drop for those 17 or over in 1985. As is to be expected, school attendance falls for all age groups between the two survey years. Clearly, decisions are being made regarding not only work activity but also school attendance.

The examination of work and school participation rates in cross-section does not shed light on the movement of individuals between the states of working and not working and on the relationship of those movements to schooling. Table 4 shows a cross-tabulation of the number of individuals in the sample according to their work and education status in 1985 and 1986. A total of 231 individuals, nearly 60 percent of the sample, did not attend school in either year. In terms of work status transitions, 57 individuals entered the work force in 1986, having not worked in 1985. Correspondingly, 65 individuals left the work force in 1986. Of those who changed their working status, most (75) did so while

Table 4. *Work and School Status of Rural Youth Ages 12 to 18, 1985 and 1986*

	Work: Both years	Work: 1985 only	Work: 1986 only	Did not work either year	Total
School					
Both years	20	47	28	44	139
1985 only	6	0	8	4	18
1986 only	1	2	0	0	3
Did not attend either year	189	16	21	5	231
Total	216	65	57	53	391

Source: World Bank data.

attending school in both years. Of those who attended school in 1985, only 8 left education to enter the work force in 1986. Clearly, participation in work does not necessarily preclude school attendance.

What has not been considered, however, is the intensity of work or schooling undertaken by individuals within the various states of work and education. The survey does contain information on the intensity of work activity, as captured by both the number of weeks worked during the year and the number of hours worked each week. Work intensity is an important issue that should be investigated in the future, but here it is just noted that, as expected, the number of hours worked each week by those not attending school is significantly higher than for those who work while attending school.

A Household-Specific Measure of Production Price Changes

One of the major features of structural adjustment is an attempt to reallocate the mix of labor—and hence output—through changes in relative prices. It is important, therefore, to identify the effect of price changes on an individual's work activity. One difficulty in estimating the response of working behavior to price incentives is the limited amount of variability observed in the two years of price data. The effect of an output price change on work incentives for a particular household, however, will depend on the importance of that particular good in the household's overall output. This means that even if all households face the same price changes for each crop, the changes in incentives are likely to be greatest for those households that specialize in the production of crops whose prices have risen most.

A great advantage of the panel is that household-specific indexes of price changes can be calculated based on the first year's production shares, thereby avoiding the possible endogeneity of current period weights. Therefore, this measure of incentive changes may vary considerably among households, at least to the extent that the composition of agricultural production differs across households and changes in prices differ across outputs. An important part of the data work underlying this study is the construction of these composite price indexes.

The household survey collects information on the quantities sold and the prices received for 24 crops. For 11 of these crops it was not possible to construct crop price indexes, either because the value of the crop in total production was rather small or because insufficient price information was available due to the fact that the price of the crop was generally reported in noncomparable and imprecisely defined units. The 13 crops used to construct the price indexes are cocoa, coffee, oil palm, plantains, fruit trees, cola nut, cotton, peanut, cassava, yam, maize, rice, and vegetables.

Average regional prices were calculated for each of five regions: north (north of Kassou Lake), southwest 1 (between Buyo Reservoir and the Guinea and Liberia borders), southwest 2 (central southern area to the southwest of Kassou Lake), southeast 1 (north of Abengourou), and southeast 2 (Abidjan hinterland

to the south of Abengourou). In calculating average prices for each crop, observations for which the quantities sold were reported in nonmetric units were excluded. Also excluded were some outlier observations, where the selection of outliers was based on a 95 percent range of data. Seasonal variation in prices could not be corrected for because the survey did not ask when during the year crops were harvested and sold.

Having calculated average regional prices, composite divisia price indexes for each household were derived using the following formula:

$$(6) \qquad P_{86h} = \sum_i w_{ih}^{85}(\ell n p_{86i} - \ell n p_{85i})$$

where w_{ih}^{85} is the value share of commodity i in total 1985 gross cash crop production for household h, and p_{86i} (p_{85i}) is the average regional price of good i in 1986 (1985). Although the value shares are household specific, the price changes vary only with region.

Two separate price indexes are calculated for each household. For the gross production index (PRODIND) the value of production used in household weighting is defined as the amount of the harvest sold, plus the value of replacement capital such as seeds, the value of any harvest given away, and the value of home consumption. For the cash crop index (CASHIND) the value of production is the net value of cash crops sold, where cash crops include coffee, cocoa, cotton, and cola nut. It was necessary to construct the CASHIND index because the observed wide dispersion in subsistence crop prices within regions makes the PRODIND index somewhat suspect. This dispersion is due, in part, to the small number of observations for some subsistence crops. Cash crop prices are generally controlled and hence more uniform, yielding a more reliable cash crop price index. Some apparent cash crops are excluded from the CASHIND index. For rubber, for example, no household in the sample had positive recorded production. Some households produce fruit products sold for cash, but these are usually recorded in noncomparable units. In calculating both indexes, only the value of primary crop production was included because of insufficient price information on transformed production and on animal products.

Table 5. *Regional Mean Price Change Indexes, 1985–86*

		Region				
Index	*Total*	*North*	*Southwest 1*	*Southwest 2*	*Southeast 1*	*Southeast 2*
Gross production index (PRODIND)	0.29 (0.42)	−0.17 (0.27)	0.23 (0.19)	0.45 (0.27)	0.76 (0.39)	0.27 (0.30)
Cash crop production index (CASHIND)	0.03 (0.02)	0.01 (0.01)	0.04 (0.02)	0.03 (0.02)	0.03 (0.03)	0.03 (0.01)

Note: Standard deviations are given in parentheses.
Source: Authors' calculations based on World Bank data.

Table 5 reports the regional mean values of the price change indexes. The prices of cash crops clearly do not change by as much as the prices for gross production. The difference between the two indexes for the northern region and those for the other regions is a consequence of the fact that the procurement price for cotton, which is the main crop within the region, was not raised by the government in 1986. The price of cocoa and coffee, which are more important in the other regions, increased by 14 and 5 percent, respectively. These figures are confirmed in Berthélemy and Bourguignon (1989).

III. Empirical Results: Incentives and Agricultural Working Behavior

Empirical studies of individual participation in education and work are usually restricted to cross-sectional analysis because of the scarcity of panel data in developing countries. This study is therefore in the almost unique position of being able to exploit the repeated observations on each individual, available across pairs of years in the Côte d'Ivoire survey. This will enable the investigation of transitions in work behavior over time as well as the assessment of the usefulness of panel data over cross-section data in such investigations. Since many decisions are based on horizons longer than two years, the models attempt to exploit the richness of the cross-section data to capture long-term influences, leaving the time series aspect to provide information on short-run transitions.

Perhaps the most natural statistical model for an analysis of working behavior over time is a discrete transition model, in which current period work status is related directly to a vector of individual characteristics, z_i, conditional on last period's work status. If there were only two initial states—described by whether an index S_i^0 equals unity if the individual works, or zero if the individual does not work—then the probability of working in the current period ($S_i^1 = 1$) would be given by

$$Pr[S_i^1 = 1] = F(z_i' \beta) Pr[S_i^0 = 1] + F(z_i' \gamma) Pr[S_i^0 = 0] \tag{7}$$

where $F(.)$ describes each discrete conditional probability, β is a vector of parameters of interest conditional on working in the initial period, and γ is a vector of parameters of interest conditional on not working in the initial period.

If β and γ are equal, there would be no need to condition on initial state. In this case, apart from time varying variables, $Pr[S_i^0 = 1] = Pr[S_i^1 = 1]$, and cross-section data alone could be used to estimate the underlying parameters. If the initial state matters, however, repeated observations over time on each individual are needed to avoid bias in the estimation of β and γ.

Referring to table 4, the individual's work status in the initial period takes one of three easily recognizable states. First, 281 young people worked in the first period. Second, of those who worked in 1985, 73 individuals also attended school in the first period. Comparing these two groups will enable us to assess whether being involved in some education in the initial period influences the

determinants of continuing to work in the next period. Finally, 84 individuals report participation only in education in the first period.

Although the purpose of conditioning on past behavior is to identify the probability of working in the second period conditional on the work status of that individual in the recent past (Heckman 1981; Nakamura and Nakamura 1985), the observance of different parameter vectors β and γ is consistent with either state dependence or unobserved heterogeneity of individuals. The experience of working in the initial period, for example, may somehow change people's constraints or incentives so that they are more likely to work in the next period. Or, different individuals may simply have different probabilities of working due to unobserved characteristics. Because of the short panel available this analysis cannot distinguish between these two explanations of behavior.

The results in this section should be interpreted as coming from a reduced form transition model, since neither the direct effects of current period schooling on work status nor the effects of current period decisions on work and schooling by other household members are estimated. Because not only the region but also the cluster to which each individual belongs can be identified, the initial empirical specification included cluster-specific dummies. These dummies could be grouped into five broad regional dummies, and indeed further grouping is possible even in a reasonably tightly parameterized setting. Furthermore all models were initially estimated on samples separated by gender and religion, but, in the results reported here, the gender and religion effects are represented by dummies.

The model estimated corresponds to equation 5, which contains several variables whose measurement posed a complex and important problem. The household price change indexes are perhaps the best illustration of this, and their definition was described fully in section II. Initial period product shares were used to weight regional price changes so that this index can be treated as exogenous in the conditional probability model. Given other factors in the model, a strong positive effect of this variable on the probability of work in the second period might be expected if incentives are having a strong effect on work activities. However, given the reduced-form nature of the labor supply equation noted above, the cash crop price variable will also capture the income effects of a change in output prices. As noted in section I, one objective of this study is to assess the relative importance of such effects.

To capture a long-run or life-cycle measure of other income and wealth consumption expenditure is used as an explanatory variable. This is likely to be a current period endogenous variable, so consumption from the initial survey is used. In addition this is restricted to cover food expenditures only, because they not only make up the largest share of a household's budget but also appear to be the best measured consumption item. These expenditures include the value of home production. The income and other household decision variables from the first period will be important in the estimation only so far as they capture longer run effects not captured by the work status variable of the initial period.

The influence of local conditions on behavior can be captured through the extensive information on local infrastructure variables and schooling availability. Other noneconomic influences can be measured through factors such as religion, age structure, and general demographic variables. Two primary measures of work status in the panel survey relate to working behavior in the past week or during the past year. The results were derived using the latter definition alone; these results were not only quantitatively similar to those using work status in the past week, but were also generally more precisely determined. This is encouraging, since it suggests the results are not unduly affected by seasonal work patterns. The results also use a similar measure for the past 12 months for schooling behavior.

A binary probit model describing the conditional probability of each individual working in 1986 was estimated by maximum likelihood for each of the three initial states, and the results are presented in table 6. In all the results a reasonably parsimonious parameterization of each model is specified. Indeed a general

Table 6. *Work Status in Period 2 Conditional on Work and School Status in Period 1*

	Status in period 1		
	Work	*Work and school*	*School only*
Constant	−2.58	−2.96	1.71
	(−2.66)	(−1.41)	(1.05)
CASHIND	15.97	38.18	−14.46
	(2.76)	(3.33)	(−1.49)
Age	0.35	0.18	−0.02
	(5.94)	(1.29)	(−0.17)
Number of household workers	−0.09	−0.17	−0.19
	(−1.77)	(−2.07)	(−2.27)
Food expenditures	−0.45	−1.03	0.92
	(−1.77)	(−1.66)	(1.40)
Muslim	0.43	1.36	
	(1.13)	(2.19)	
Male			−0.53
			(−1.16)
Regional dummies			
Southwest 1	−2.00	n.a.	−0.13
	(−3.38)	n.a.	(−0.14)
Southwest 2	−2.35	n.a.	−0.94
	(−4.56)	n.a.	(−1.09)
Southeast 1	−1.41	0.04	−0.65
	(−2.48)	(0.07)	(−0.72)
Southeast 2	−1.26	1.22	−1.88
	(−2.11)	(2.38)	(−2.01)
Number of observations	281	73	84

n.a. Not applicable.

Note: The dependent variable is binary and equals 1 if the individual worked in period 2; it equals 0 otherwise. The *t* statistics are in parentheses.

Source: Authors' calculations based on World Bank data.

feature of the results was the dominating importance of the cash crop price effects measured by CASHIND over the general index PRODIND, so only CASHIND results are given. The inferiority of PRODIND as an explanatory variable probably reflects problems in measuring the prices of noncash crops. A full set of results, programs used, and detailed data transformations are available from the authors.

Starting with individuals who worked in the initial period, the individual's age and the number of other household workers in the initial period were found to be of some importance. A variable measuring land size was also included in preliminary models, but this was never found to play a role once regional effects were allowed for. Food expenditures in the initial period were added to capture any longer-run income effects not reflected in the discrete state conditioning variable. Its negative coefficient confirms that income affects work status. It was thought that the individual's religion could be important, but its effect was small and insignificant. Location variables could play an important role in shaping both preferences and constraints. The regional dummies are clearly seen to be important, although some further grouping looks possible. Finally, several infrastructure variables and a gender variable turned out to be of little significance for this group.

The issue of state dependence can be examined by comparing the behavior of the subset of 73 individuals from this group who had not only worked in the first period but had also been in school. The results for this subgroup are reported in the middle column of table 6. A similar overall pattern is observed, but an even stronger cash crop price effect appears. In this smaller sample both southwest regional dummies were grouped with the base region (north).

Finally, the results for the group of 84 individuals who were wholly engaged in education in the first period contrast distinctly with those for the first two groups. Most notably, the cash crop price effect is negative (although insignificant). Those who are not already working are less likely to respond to favorable movements in the prices of cash crops by entering the work force. In fact the probability of working declines with increases in the price index, which indicates that the income effect of price movements dominates for this group.

IV. Conclusions

This article has been concerned with modeling the participation in work of young people in rural Côte d'Ivoire. Since many young people combine work and school, the situation is more complex than a simple choice between the two. The econometric analysis has attempted to identify the factors that determine the choice to work. The panel aspect of the data allowed the choice made in one year to depend on the individual's choice in the previous year.

The analysis of the decision to participate in the labor force shows that, among other factors, the price of agricultural output generally has a positive incentive effect, but this result depends heavily on the employment and educa-

tion state in which the individual is placed in the first period. This is of interest not only because it suggests that, for those already working, the substitution effect of the change in wages outweighs the income effect, but also because it shows that the construction of household-specific price indexes can help to explain labor supply behavior with panel data sets of this type.

Using food consumption to replace income as a measure of welfare was more disappointing. In contrast to most studies of labor supply, this measure does not appear to exert a strong influence. This could be a result of the inaccurate measurement of food consumption. Although the food consumption data were better than other available consumption measures, brief analysis of the data showed both an unexpectedly large inequality in food consumption per capita and implausibly large year-to-year variation in food consumption per capita for many of the panel households. Further research to find a more accurate measure of consumption is warranted.

The results of this study suggest that the CILSS data can be used to analyze labor supply behavior and that their panel aspect is of considerable value. The analysis also suggests several extensions. A preliminary examination of hours of work and their relation to schooling suggests that the modeling and estimation of working hours could be very important in estimating the effect of structural adjustment on overall labor supply. Moreover the analysis presented here and the proposed analysis of working hours should be extended to other age groups, and the decision to pursue education should also be explicitly addressed.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Berthélemy, Jean Claude, and François Bourguignon. 1989. "Growth and Crisis in Côte d'Ivoire." Département et Laboratoire d'Economie Théorique et Appliquée, Paris. Processed.

Deaton, Angus, and Dwayne Benjamin. 1987. "Household Surveys and Policy Reform: Cocoa and Coffee in the Côte d'Ivoire." Research Program in Development Studies Discussion Paper 134, Princeton University. Processed.

Glewwe, Paul, and Dennis de Tray. 1988. *The Poor during Adjustment: A Case Study of Côte d'Ivoire*. LSMS Working Paper 47. Washington, D.C.: World Bank.

Heckman, James J. 1981. "Heterogeneity and State Dependence." In Sherwin Rosen, ed., *Studies in Labor Markets*. University of Chicago Press.

Nakamura, Alice, and Masao Nakamura. 1985. "Dynamic Models of the Labor Force Behavior of Married Women Which Can Be Estimated Using Limited Amounts of Past Information." *Journal of Econometrics* 27(3):273–98.

Singh, Inderjit, Lyn Squire, and John A. Strauss, eds. 1986. *Agricultural Household Models: Extensions, Applications, and Policy*. Baltimore, Md.: Johns Hopkins University Press.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 155-169

Social Security and Private Transfers in Developing Countries: The Case of Peru

Donald Cox and Emmanuel Jimenez

Do social security systems "crowd out" private transfers from younger to older generations? This question has generated much theoretical discussion, but little empirical work exists to confirm or refute this crowding-out hypothesis. We investigate the connection between social security and private transfers in Peru, using the Peruvian Living Standards Survey, and find that private transfers from young to old would have been nearly 20 percent higher without social security benefits. This indicates that the Peruvian social security system is less effective at delivering benefits to the elderly than a simple assignment of government expenditures would suggest. Social security's displacement of private transfers, while significant, is less than that predicted by models with widespread altruistic transfers.

How effective are social security retirement benefits at raising the incomes of older households in developing countries? On the surface it seems that a "pay as you go" social security system, which taxes working households and distributes the proceeds to those who have retired, should benefit older households at the expense of younger ones. But this simple assessment could be inaccurate because of the effect of public transfers on informal private transfers among families.

Consider the hypothetical case of an elderly couple supported by their children. If a social security program is created that taxes the children and channels the funds to the elderly household, the children may reduce their private transfers because their social security contributions accomplish what they once did informally. Such a response would weaken the impact of social security on income distribution. Indeed some U.S. researchers, notably Becker (1974) and

Donald Cox is on the faculty of the Department of Economics at Boston College and a consultant to the World Bank. Emmanuel Jimenez is with the Population and Human Resources Department of the World Bank. The work on this article was supported in part by RPO 674-49, granted to the Public Economics Division by the World Bank's Research Support Budget. The authors wish to thank the government of Peru and the Bank's Welfare and Human Resources Division for providing access to data and the three anonymous referees and participants at a World Bank seminar for comments on an earlier draft. Jorge Castillo, Kalpana Mehra, and Reza Firuzabadi were invaluable in setting up workable data files. Fiona Mackintosh provided editorial assistance.

Barro (1974), argue that private transfers could render social security programs completely ineffective.

Are such considerations important in developing countries that are considering reforms in their social security system? Private transfers are a major component of household income and expenditure in virtually all of the developing countries surveyed by Rempel and Lobdell (1978) and Cox and Jimenez (1990), and much of the transfer income flows from children to parents. For example, more than a quarter of the private transfers in a sample of Kenyan households were given to parents from their children (Knowles and Anker 1981). Butz and Stan (1982) found significant transfers from young to old in Malaysia, as did Ravallion and Dearden (1988) in Java. In fact some researchers argue that, in developing countries, security for one's old age is the most important motive for having children, so that the prevalence of transfers from young to old should not be surprising (Leibenstein 1975). (For alternative views on this controversial hypothesis, see Vlassoff and Vlassoff 1980 and Nugent 1985.)

While the potential linkage between social security and private transfers has obvious significance for policy, very little empirical work has been done on this issue. Part of the problem is the lack of adequate data; data sets containing both private transfer and social security information are scarce. Many countries for which information about private transfers exists have no social security programs or devote only a negligible fraction of their budgets to public transfers of income. Data are available for some industrialized countries, such as the United States, but using these data to make inferences about developing countries would likely be misleading, given fundamental differences in the size of the social security programs and the patterns of intrafamily transfers. In fact, in the United States, private transfers from young to old are negligible (Cox and Raines 1985), which would make it difficult to gauge any connection between social security and private transfers. Furthermore comparisons among countries with data for private transfers are not feasible because of disparities both in survey definitions of private transfers and in sampling techniques. Some surveys, for example, count transfers in kind, whereas others do not, and some surveys focus on low-income subsamples, whereas others use representative cross-sectional samples.

A new data set, the Peruvian Living Standards Survey (PLSS), is used here to assess the link between social security and private transfers. Peru is particularly well suited as a case study of the effects of social security on private transfers. The country is an ideal balance between the extremes of very poor nations, like Kenya, which have significant private transfers but little public transfer spending, and the United States, which has a large social security system but negligible private transfers from young to old. In Peru a significant fraction of private transfers flows from young to old, and the social security program has expenditures equal to more than 3 percent of the gross domestic product (Suarez-Berenguela 1987). Furthermore the PLSS contains the labor-market and demo-

graphic information that must be controlled for in order to isolate social security's impact on private transfers.

The sections that follow review theories about private transfers in more detail and describe the Peruvian social security system. The empirical sections provide an overview of private transfers in Peru and measure the extent to which social security supplants private transfers.

I. Theories about Private Transfers

The response of private transfers to social security depends on motives for giving. Economic models generally predict that if motives are purely altruistic and private transfers are widespread, social security will completely crowd out private transfers. If transfers are at least partly motivated by self-interest, complete crowding out may not occur, even if altruistic motives are also present.

Modern analysis of the connection between social security and private transfers began with the seminal work of Barro (1974), which uses the altruistic framework for private transfers that was advanced by Becker (1974). In this model the utility of the child, U_k, depends on own consumption, C_k, and the utility of the parent, U_p. The parent's utility, in turn, depends on own consumption, C_p, so that

$$U_k = U_k[C_k, U_p(C_p)]. \qquad (1)$$

Both child and parent are subject to budget constraints such that the child's consumption must equal own income, I_k, net of any transfers, T, to the parent. The parent's consumption must equal own income, I_p, plus transfers received from the child.

The maximization of equation 1, assuming an interior solution, implies that transfers will be given to achieve the optimal consumption of each family member. Each member's consumption depends only on aggregate family income, $I_k + I_p$, and not on the distribution of its components. A social security program that forces a transfer from child to parent but leaves aggregate family income unchanged will have no effect on either family member's consumption. The child will reduce private transfers by the exact amount of the forced public transfer to maintain own consumption and that of the parent's at their previous levels. In summary the theory of altruistically motivated transfers predicts that transfers through social security merely supplant (that is, completely crowd out) private ones, leaving individual consumption and well-being unchanged.

There are several reasons why complete crowding out may not occur. Although it is technically possible for crowding out to be incomplete in a purely altruistic model if there are corner solutions, most models achieve incomplete crowding out because of the introduction of self-interest as a motive for giving. As an illustration, consider a model based on pure self-interest, in which neither parent nor child cares about the other. Suppose it is difficult to use capital

markets to borrow or save, so that parents lend to their children, who repay when they reach middle age.[1] Changes in social security would influence self-interested lending behavior only to the extent that the current income and lifetime wealth of generation members are affected. Now suppose social security taxes and benefits increase in a way that leaves the lifetime wealth of the younger generation unchanged but increases the lifetime wealth of the older one. The current income of the young would be depressed, but current desired consumption would remain unchanged. If the young are liquidity-constrained, they would borrow more from the older generation, which would imply higher repayments to the older generation later on. The enhanced lifetime wealth of members of the older generation would reinforce this effect, because it would put them in a better position to lend by lessening any credit constraints they would face. The expansion of social security would thus result in higher future transfers from young to old, contrary to the prediction of the altruistic model.

This example is somewhat contrived, as it is designed only to illustrate the possibility that self-interest can generate a different outcome than does altruism. It is also possible to construct examples in which the qualitative predictions for self-interested motives are similar to those of altruistic models. For example, a common theme in the development literature is that private transfers from young to old might be repayments for past assistance or insurance premiums against income shortfalls (Rempel and Lobdell 1978). If an expansion in social security reduces the lifetime wealth of younger generations, such transfers from young to old in the form of loan repayments can decline because current consumption and borrowing will have fallen. The exact crowding out of private transfers by public ones, however, which is the primary implication of the altruism model, does not generally occur if transfers are motivated by self-interest.

An alternative specification contains a mixture of altruism and self-interest. Both Lucas and Stark (1985) and Kotlikoff and Spivak (1981) posit that a combination of self-interest and altruism governs private transfer behavior. Self-interest prompts households to enter an agreement to, say, lend or insure, but altruism creates the mutual trust necessary to circumvent the moral hazard inherent in such agreements. These models are attractive because they explain why households motivated by self-interest are induced to repay loans.

This eclectic specification of transfer motives will not result in complete crowding out except in special cases. Even if altruism of the Barro-Becker type is combined with self-interest, only one transfer motive can determine the comparative statics results in a given instance, depending on which motive is operative at the margin (Cox 1987). For example, suppose the transfer of a household's

1. There is evidence that private transfers respond to capital market imperfections (Cox 1990). If transfers are used to smooth consumption, transfers and earnings should move inversely over the life cycle. Elsewhere (Cox and Jimenez 1990) it is shown that the U-shaped age pattern for the probability of transfer receipt is almost the mirror image of the age-earnings profile, which supports the idea that private transfers respond to capital market imperfections.

last bit of income is motivated by self-interest and purchases some service from the recipient. In this instance marginal changes in pretransfer incomes induced by modifications in social security will not produce the Barro-Becker crowding-out results even though the donor has altruistic feelings, because altruistic motives are not operative at the margin. The donor is made happier if the recipient receives a windfall income gain but is not willing to make a transfer to raise the recipient's utility. Furthermore the Barro-Becker model employs an important assumption about the bargaining framework between donor and recipient—the donor dominates. Alternative frameworks, such as Nash bargaining, do not result in complete crowding out (McElroy and Horney 1981; Cox 1987; Kotlikoff, Razin, and Rosenthal 1990).

A final motive for private transfers is one that is often neglected by economists but emphasized by other social scientists: the existence of social norms, such as guilt feelings, pressure family members into helping other members in need (see Becker 1988). In many societies such feelings can extend to include distant relatives or village members. Although precise models about how these norms evolve are not available, they nonetheless are likely to be important in explaining why even individuals motivated by self-interest transfer resources to one another. But social norms need not be modeled similarly to pure altruism. For example, social norms can be interpreted so that the perception (by others in the community) of giving, rather than its outcome, determines the donor's utility. In this case complete crowding out need not occur.

II. The Social Security System in Peru

Peru's social security system (the Spanish abbreviation is IPSS), which began in the early 1930s as a risk-sharing plan for blue-collar workers, covered about 40 percent of the economically active population by 1985.[2] Participation is mandatory for all employees of public or private firms and cooperatives; self-employed workers have an option to participate in the system. Because of the focus on wages and salaries, coverage is concentrated in the high-income urban and formal sectors of the economy: white-collar workers, blue-collar workers in large firms, and the military.

The main sources of funds are payroll contributions by insured persons and their employers. Since 1987 covered workers have been expected to contribute 5 percent of their salaries to IPSS, whereas employers contribute 14 percent of wages and salaries paid, plus 1 to 12 percent for work injury compensation, depending on the risk of the occupation (U.S. Social Security Administration 1987). Self-employed workers contribute 15 percent of their monthly income, but the income basis for their contributions is subject to maximum and minimum levels. The government contributes to the system primarily as an employer.

2. This section draws heavily on Mesa-Lago's (1985) and Suarez-Berenguela's (1988) description of the IPSS.

Peru's social security system provides pension and health care benefits, with ne-half of the employee and employer tax revenues earmarked for pensions and ie other half funding medical and maternity expenses. Medical and maternity enefits comprise a little under two-thirds of IPSS's total expenditures, with ension funds used to replenish the depleted health budget. In 1983 the system ; a whole ran a deficit equal to 6.7 percent of total receipts (International abour Organization 1987).

III. Description of the Data

The data set used in this study is the Peruvian Living Standards Survey (PLSS), nducted by the World Bank in conjunction with the Peruvian Instituto Nacio-l de Estadistica. The PLSS contains socioeconomic information for a sample of 109 households surveyed between June 1985 and July 1986.

The household is the unit of observation for the analysis below. This limits e analysis to interhousehold transfers, but this might not be a strong limita-n. Many intrahousehold allocation models, such as Rosenzweig (1986), as-me that households reallocate resources to maximize a unified welfare func-n. Also these models predict that households will at least partly offset distribution effects, as found here in the analysis of interhousehold transfers.

Households missing data on private transfers, age, education or gender of usehold head, parental schooling, illness, household size, or consumption ere deleted, as were households with no indicator of urban or rural residence. iese deletions reduced the sample size to 4,184. Furthermore, because social curity is primarily an urban phenomenon, with urban households accounting r 85 percent of social security coverage, rural households were excluded, ducing the sample to 2,241.

Survey respondents were asked the following: "Has any member of your usehold sent money or goods—regularly or irregularly—to persons who are t members of your household during the past three months?" Respondents ere also asked to report the relationship between the recipients and the house-ld head (for example, son, parent) and the value, in intis, of cash and transfers kind given in the preceding three months. The same questions were asked for erhousehold transfers received.

Of the 2,241 households in the sample, 760 (or 33.9 percent) reported giving rivate transfer to another household, and 723 (32.3 percent) reported receiv-g a private transfer from another household. Two hundred eighty-two (12.6 rcent) reported both giving and receiving a transfer, while 1,040 (46.4 per-nt) neither gave nor received.

Because some households both gave and received transfers, this analysis fo-ses on net transfers received (transfers received minus transfers given) and net insfers given (transfers given minus transfers received). A household is defined a net transfer recipient if gross transfers received exceed gross transfers given. household is a net transfer donor if gross transfers given exceed gross transfers

received. The sample contained 574 net transfer recipients (25.6 percent of the sample), 613 net transfer donors (27.4 percent), and 1,054 households for which net transfers were zero (47 percent). Because 282 households both gave and received a transfer, some givers are included in the "net transfer recipients" category, and some recipients are included in the "net transfer donors" category. For simplicity, throughout the rest of the article net transfer recipients will be referred to as "recipients" and net transfer donors as "donors" or "givers."

The average of net transfers received for the entire sample is 77.70 intis monthly—4 percent of average total monthly income. To put this figure in perspective, this is roughly two and a half times the average social security pension income. The average net transfer given is 71.05 intis. Curiously, there seems to be no tendency to exaggerate transfers given or to underreport transfers received. If anything, any reporting bias appears to have gone in the other direction. This finding runs counter to the evidence reported in Cox and Raines (1985), in which reported transfers given often exceeded transfers received for a cross-section of the U.S. population. The average net transfer receipt among Peruvian recipients is 303.36 intis—22 percent of the average income of recipients before private transfers. The average net transfer given among donors is 259.76 intis—11.8 percent of average donor income. Private transfers are thus nontrivial in both number and magnitude.

Survey respondents were asked to report the main sources of transfers received and the destinations of transfers given. A summary of the sources of transfers received is shown in table 1. Most of the transfers occur between parents and children. After these two categories, the most significant is that of "other relatives," who are the source of 29 percent of the transfers received, although less than a fifth of total intis received. Very few transfers come from grandchildren or spouses, but nonrelatives account for a significant minority of transfers. Because this analysis focuses on interhousehold transfers, interspousal transfers are not considered unless the spouses are living in separate households. The category that is the focus of the empirical work below, transfers from children to parents, is the largest source of transfers, accounting for 32.5 percent of all private transfer income.

Table 1. *Sources of Private Transfer Income in Peru, 1985–86*

Source	*Number of transfers*	*Percentage of transfers*	*Average transfer (intis)*	*Percentage of total intis received*
Parents	191	27.6	248.3	26.1
Other relatives	204	29.4	176.5	19.8
Children	182	26.3	324.9	32.5
Grandchildren	9	1.3	101.3	0.5
Spouse	26	3.8	477.3	6.8
Nonrelatives	81	11.7	320.4	14.3

Note: Because households can receive from more than one source, the number of transfers reported here exceeds the number of households receiving net transfers.

Source: Calculations from the Peruvian Living Standards Survey, 1985–86.

IV. What Determines Transfer Behavior?

This section investigates empirically the question posed at the outset: Does social security crowd out private transfers to parents from children? There are two issues to explore. First, what is the connection between social security and the occurrence of a transfer? Second, when a transfer occurs, how does social security affect the amount?

Probit analysis is used to analyze the first question. The analysis focuses on the sample of 1,121 households whose head is age 45 or over. Indexing households by h and adding a normally distributed stochastic component, the latent variable that determines the transfer decision is expressed as

$$(2) \qquad \begin{aligned} t_h &= a_0 + aI_h + bX_h - \epsilon_h, \\ \text{and } T_h &> 0 \qquad \text{iff } t_h > 0, \\ T_h &= 0 \qquad \text{otherwise.} \end{aligned}$$

When the latent variable t_h crosses the threshold 0, transfers, T_h become positive. Otherwise, they are zero. The vector I_h is a set of three variables having to do with pre-private-transfer income: a dummy variable indicating whether the household receives social security income; the amount of social security income; and the amount of non-social-security income, which includes earnings, financial income, rental income, and income from various other sources, such as gambling.

The vector X_h contains education and demographic variables that may affect the incidence of transfers, including dummies for whether anyone in the household has been ill during the past four weeks or unemployed during the past 12 months. To measure additional household resources, a dummy indicating whether the household head is a homeowner is included. The vector also contains dummies for educational level and a quadratic in age. Previous studies of transfer behavior indicate that the gender of the head of household is an important determinant of transfer behavior. Evidence from developing countries (for example, Lucas and Stark 1985 for Botswana and Kaufmann and Lindauer 1986 for El Salvador) indicate a positive relationship between transfers and female status, and a similar pattern has been found for the United States (Cox 1987). A dummy variable indicating whether the head of household is female is therefore included. Finally, X_h contains dummy variables for marital status, the number of children under age 30 living outside the household, whether there are no children under 30 living outside the household, and household size.

Heckman's (1979) generalized tobit analysis is used to examine the effect of social security on transfer amounts. The estimating equation for transfer amounts is given by:

$$(3) \qquad T_h = c_0 + cI_h + dX_h + E(\eta_h \mid T_h > 0)$$

where η_h is a random error component whose expectation is conditional on the

observation of a positive transfer. The composition of the I_b and X_b vectors here is similar to that used in the probit equation, except that age enters linearly.[3]

Table 2 contains the probit and generalized tobit estimates for transfers received. The dummy variable for whether the household receives income from social security is negative, large, and statistically significant at the 0.01 level. At sample means, having income from social security reduces the probability of transfer receipt by 8 percentage points. This is the main finding of the article, and it forms the basis of our assessment of the impact of social security on transfers from young to old. First, however, we discuss the other findings of interest from table 2 as well as the determinants of giving.

Neither the level of social security income nor the level of non-social-security income (labeled simply "income" in table 2) is statistically significant in determining transfer receipt, and their coefficients are very small. At sample means, a 1,000-inti increase in income reduces the probability of receiving a transfer by only half a percentage point. The most important income measure affecting the receipt of a transfer is the dummy for the receipt of social security income.

One possible explanation for the pronounced nonlinear effect of social security is that the dummy for the receipt of benefits captures the effects of potentially weaker family ties among those covered by IPSS. This issue was explored further by searching the PLSS data for family-tie indicators to add to the list of regressors already used. Four variables that could potentially serve as additional measures of the strength of family ties were a dummy indicating whether the main respondent was born in a rural area, the number of times the respondent migrated, the number of years the respondent had lived in the area, and occupation. These variables were added to the transfer probit and found not to exert a statistically significant effect on transfer incidence. The rural-birth dummy was positive and on the margin of statistical significance, but including these variables had a negligible effect on the coefficient of the social security dummy.

The most plausible explanation for the pronounced discontinuity in social security's effect stems from the inclusion of health benefits in social security payments. Because people collecting social security benefits also enjoy better health coverage and because private transfers respond to health risks, a sharply discontinuous effect on the incidence of private transfers is not all that surprising.

The probability of receiving a transfer rises with age up to age 81 and then declines. At the sample mean (age 57), being one year older raises the probability of transfer receipt by 2.4 percentage points. The probit results indicate that transfers are targeted toward homeowners, the unemployed, and house-

3. The probit equation used to generate the inverse Mills ratio in the generalized tobit contains the same vector of explanatory variables as does the probit in equation 2, plus additional terms. Income, age, marital status, female status, illness, and unemployment are entered interactively. The extra variables are used to identify the generalized tobit. Results from the expanded probit equation are available on request.

Table 2. *Probit and Generalized Tobit Estimates of Transfers Received by Parents from Children in Peru, 1985–86*

Variable	Probit: *Coefficient*	Probit: *Asymptotic t-value*	Generalized tobit: *Coefficient*	Generalized tobit: *t-value*
Income variables				
Income	-0.290×10^{-4}	−1.14	0.029	1.16
Social security recipient	−0.565	−3.26	129.695	0.65
Social security income	0.148×10^{-3}	0.70	−0.295	−1.29
Education dummy variables[a]				
Initial	−3.377	−0.13	n.a.	n.a.
Primary	0.271	1.60	15.850	0.11
Secondary	0.259	1.32	443.218	2.53
Technical	−0.172	−0.43	780.340	1.58
Postsecondary	−0.131	−0.30	36.904	0.08
University	0.058	0.22	427.192	1.60
Household characteristics				
Age	0.204	3.46	6.605	0.78
Age squared	−0.001	−2.72	n.a.	n.a.
Unmarried	0.057	0.29	−195.191	−1.05
Female	0.454	3.49	355.273	2.46
Ill	0.091	0.86	−251.910	−2.33
Unemployed	0.393	3.50	−35.258	−0.30
Homeowner	0.273	2.34	−68.908	−0.58
Number of young[b]	−0.097	−0.51	−95.663	−0.50
Young in school[c]	0.146	0.76	100.045	0.54
No young	−0.272	−1.86	149.073	0.96
Household size	−0.008	−0.42	45.155	2.15
Constant	−9.029	−4.83	−436.435	−0.49
Inverse Mill's ratio	n.a.	n.a.	22.545	0.10
Recipients	175		175	
Observations	1,121		175	
R-squared	n.a.		0.22	
Dependent variable mean	0.16		330.29	

Note: In the probit analysis, the dependent variable is a transfer receipt (transfer receipt = 1 if the transfer is received from a child, 0 if otherwise). In the generalized tobit analysis, the dependent variable is the net transfer amount received. The sample consists of households whose heads are age 45 or over.

n.a. Not applicable.

a. The reference category is no education.

b. Number of children under age 30 living outside the household.

c. Number of children who are in school living outside the household.

Source: Calculations from the Peruvian Living Standards Survey, 1985–86.

holds headed by females. Education is not significantly related to transfer receipt.

For the PLSS, households were asked for information on children under the age of 30. Having no children under 30 reduces the probability of transfer receipt, but the other child-related coefficients are not significantly different from zero. More than half of the transfer recipients have no young children outside the household, so they must be receiving transfers from children over 30.

The generalized tobit estimates are also given in table 2. Transfer amounts are higher for those with secondary, technical, and university education (the reference category is "no education") and lower for those who are ill. Amounts are also higher for female-headed households and larger households. None of the income measures in the generalized tobit is statistically significant.

Estimated equations for transfers given by younger to older generations are presented in table 3. The sample is restricted to the 1,875 households having at least one living parent residing outside the household. The probability of giving a transfer rises with earnings, but at a diminishing rate. At the sample mean a 1,000-inti increase in earnings raises the probability of giving a transfer by 2.2 percentage points. The probability of giving also increases with other income. Education has a negligible effect on the probability of giving. The probability of giving a transfer rises with age up to age 38, then falls. Female-headed households and the unemployed are less likely to give, but the ill are more likely to do so.

The education levels of parents and of the spouse's parents are included as indicators of the pretransfer resources of members of the older generation. These indicators show a mixed pattern for transfers given. For example, the probability of giving is inversely related to the mother's education but positively related to the spouse's mother's education.

Generalized tobit estimation yields few significant variables influencing transfer amounts given. Transfers given increase with donor earnings, but, as in the probit equation for giving, education has a negligible effect on donations. The unmarried give more, and amounts given increase with the in-laws' education.

V. The Effects of Social Security on Transfers from Young to Old

Using the estimates in tables 2 and 3, one can determine how private transfers would change if social security pension benefits were taken away. Such a determination can be made by comparing predicted transfers with and without social security. Predicted transfers are the product of the predicted transfer probability and the predicted amount. Because probit analysis has been used, the predicted transfer probability is given by the normal cumulative distribution function evaluated at $\hat{a}_0 + \hat{a}\bar{I}_h + \hat{b}\bar{X}_h$, where $\hat{a}_0$, $\hat{a}$, and $\hat{b}$ are estimated coefficients and $\bar{I}_h$ and $\bar{X}_h$ are means of the vectors of explanatory variables. Predicted transfer amounts are the products of the estimated coefficient vector and the vector of explanatory variables evaluated at sample means:

$$\text{(4)} \qquad \text{Predicted transfers} = F(\hat{a}_0 + \hat{a}\bar{I}_h + \hat{b}\bar{X}_h) \times (\hat{c}_0 + \hat{c}\bar{I}_h + \hat{d}\bar{X}_h)$$

where F is the normal cumulative distribution function. The sample-selection term and its estimated coefficient are included in predicted values for transfer amounts. (For a discussion of prediction in tobit models, see Maddalla 1984, chap. 6.) Predicted transfers without social security are given by equation 4 with social-security-related elements of the vector I_h set to zero. For transfers received

Table 3. *Probit and Generalized Tobit Estimates of Transfers Given by Children to Parents in Peru, 1985–86*

	Probit		*Generalized tobit*	
Variable	*Coefficient*	*Asymptotic t-value*	*Coefficient*	*t-value*
Income variables				
Earnings	0.128×10^{-3}	3.38	0.022	5.56
Earnings squared	-0.677×10^{-8}	−2.33	n.a.	n.a.
Other income	0.381×10^{-3}	1.99	−0.052	−1.06
Education dummies[a]				
Initial	0.820	1.45	−66.177	−0.51
Primary	−0.116	0.44	−24.783	−0.36
Secondary	0.275	1.01	−31.490	−0.45
Technical	0.338	1.12	−43.154	−0.57
Postsecondary	0.296	0.94	−48.061	−0.63
University	0.289	0.99	−18.952	−0.26
Household characteristics				
Age	0.203	2.39	0.202	0.21
Age squared	−0.004	−2.16	n.a.	n.a.
Age cubed	0.234×10^{-4}	1.77	n.a.	n.a.
Unmarried	0.242	1.16	219.491	4.78
Female	−0.367	−2.57	38.067	1.07
Ill	0.179	2.41	−21.979	−1.31
Unemployed	−0.136	−1.28	21.888	0.87
Homeowner	−0.105	−1.36	15.444	0.95
Household size	−0.048	−2.52	2.026	0.47
Number of living parents[b]	0.092	−2.22	−2.703	−0.30
Education of parents				
Father	0.016	1.21	−3.984	−1.39
Mother	−0.023	−1.47	2.344	0.69
Spouse's father	−0.022	−1.77	6.164	2.22
Spouse's mother	0.029	1.99	5.735	1.79
Constant	−4.361	−3.48	157.314	1.38
Inverse Mill's ratio	n.a.	n.a.	−90.513	−2.00
Donors	305		305	
Observations	1,875		305	
R-squared	n.a.		0.30	
Dependent variable mean	0.16		106.51	

Note: In the probit analysis, the dependent variable is a transfer given (transfer given = 1 if the transfer is given to a parent, 0 if otherwise). In the generalized tobit analysis, the dependent variable is the net transfer amount given. The sample consists of households whose heads or spouses have at least one parent living.

n.a. Not applicable.

a. The reference category is no education.

b. Includes head's and spouse's parents.

Source: Calculations from the Peruvian Living Standards Survey, 1985–86.

by old from young, the actual mean is 51.56 intis, the predicted mean is 52.13 intis, and the predicted mean without social security income is 59.20 intis. Removing social security causes a 7.07-inti increase in predicted private transfers received by older generations from younger ones. Thus, without social security, transfers per household from young to old in the sample would rise

from 51.56 intis to 58.63 intis. The boost in transfers amounts to 12.1 percent (7.07/58.63) of pre-social-security transfers.

Because donors cannot be matched with recipients in the PLSS, the 12.1 percent figure does not take into account how the removal of social security would affect donor behavior. Taking away social security increases the disposable earnings of potential donors, and, as indicated in table 3, giving is positively related to earnings. If the social security tax is fully shifted to workers, they would pay both the employee and employer portions of the assessment, or a tax rate of 19 percent. Removing the tax would raise disposable earnings by 23 percent (0.19/(1-0.19). The estimates in table 3 imply that the increase in disposable earnings from removing the tax would generate a predicted increase in private giving of 2.92 intis. Adding together the increase in transfers that would be received with the increase in transfers that would be given as a result of the tax-reduction effect yields an increase in total transfers of 9.99 intis. Without social security, then, total transfers per household would be 61.55 intis. The estimated reduction in transfers due to social security would be 16.2 percent (9.99/61.55). Another way to express this result is that without social security, private transfers from young to old would have been 19.4 percent larger (9.99/51.56).

These simulation results indicate that social security has a significant impact on private transfers, but its effect is considerably less than the Barro-Becker prediction of complete crowding out. The estimates in tables 2 and 3 can be interpreted in the light of alternative transfer motives. Some findings in table 2 are inconsistent with the strict Barro-Becker altruism motive for private transfers. The altruism model predicts a large reduction in private transfers, with increases in the pretransfer income of recipients. Instead, two of the income measures in the generalized tobit equation for transfer receipts are positively related to transfer amounts, although the coefficient on the amount of social security income is negative and large, which is consistent with the altruism hypothesis (although none of the coefficients is precisely estimated). One intriguing explanation is that donors respond more strongly to changes in the recipent's income if the changes are beyond the recipient's control. Presumably recipients have more control over their earnings than they have over public transfer receipts.

Although the probit analysis yields a negative coefficient on the receipt of social security income, which suggests altruism, probit results are a less discriminating test of transfer motives than are tobit results. The reason is that a negative coefficient on social security in the probit equation is consistent with exchange-motivated transfers as well, because a parent who expects to receive social security payments has less incentive to enter into exchange relationships with children, such as intergenerational lending or insurance. The same argument applies to the positive coefficient on unemployment status in the probit for transfer receipt. The positive relation between transfers and unemployment may be due to either altruistic or exchange motives.

One problem with making inferences about transfer motives from the equations that estimate transfer amounts, however, is the lack of information about potential donor income. For example, a possible reason why the coefficient on recipient income is positive is that the characteristics of potential donors are omitted from the equation. Positive correlation between recipient and donor incomes imparts an upward bias to the income coefficients in table 2.

Although far from conclusive, some of the evidence is consistent with interfamily exchange. For example, females are much more likely to receive transfers in Peru, which is consistent with evidence from other countries (Lucas and Stark 1985 for Botswana; Kaufmann and Lindauer 1986 for El Salvador; Cox 1987 or the United States). Part of the reason for the gender difference in transfers eceived may be attributable to the possibility that women are more involved in ie interfamily exchange of services in kind. "Altruistic" explanations, such as omen being compensated for deficiencies in earning potential, are less convincig, because the estimates control for income. Part of the gender difference could lso be caused by gender differences in life expectancy, resulting in a targeting of rivate transfers to widows.

VI. Conclusion

Like many other countries, Peru established a social security system in part to isure the well-being of the elderly. But who really benefits from the system? ssignment of public benefits does not occur in a vacuum—the private sector sponds by withdrawing some of its support. The extent of the decline in rivate transfers is an empirical question addressed in this article. Peruvian ouseholds whose heads receive social security pension income are less likely to ceive private transfers from their children. Without social security, the amount f private transfers from young to old would have been almost 20 percent larger. hus, an evaluation of policy effects should take private behavior into account.

References

irro, Robert J. 1974. "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82 (6, November/December): 1095–17.

cker, Gary S. 1974. "A Theory of Social Interactions." *Journal of Political Economy* 82 (6, November/December): 1063–93.

——. 1988. "Family Economics and Macro Behavior." *American Economic Review* 78 (1, March): 1–13.

tz, William P., and Peter J. E. Stan. 1982. "Interhousehold Transfers and Household Structure in Malaysia." *Population and Development Review* 8 (Supplement): 92–115.

ix, Donald. 1987. "Motives for Private Income Transfers." *Journal of Political Economy* 95 (3, June): 508–46.

——. 1990. "Intergenerational Transfers and Liquidity Constraints." *Quarterly Journal of Economics* 105 (1, February): 187–217.

Cox, Donald, and Emmanuel Jimenez. 1990. "Social Objectives through Private Transfers: A Review." *The World Bank Research Observer* 5 (2): 205–18.

Cox, Donald, and Fredric Raines. 1985. "Interfamily Transfers and Income Redistribution." In Martin David and Timothy Smeeding, eds., *Horizontal Equity, Uncertainty, and Measures of Well-Being*. University of Chicago Press.

Heckman, James J. 1979. "Selection Bias as Specification Error." *Econometrica* 47 (January): 153–61.

International Labour Organization. 1987. *The Cost of Social Security, 1981–83*. Geneva.

Kaufmann, Daniel, and David L. Lindauer. 1986. "A Model of Income Transfers for the Urban Poor." *Journal of Development Economics* 22 (2, July/August): 337–50.

Knowles, James C., and Richard Anker. 1981. "An Analysis of Income Transfers in a Developing Country." *Journal of Development Economics* 8 (2, April): 205–26.

Kotlikoff, Laurence J., Assaf Razin, and Robert W. Rosenthal. 1990. "A Strategic Altruism Model in Which Ricardian Equivalence Does Not Hold." *Economic Journal* 100 (403, December): 1261–68.

Kotlikoff, Laurence J., and Avia Spivak. 1981. "The Family as an Incomplete Annuities Market." *Journal of Political Economy* 89 (2, April): 372–91.

Leibenstein, Harvey. 1975. "The Economic Theory of Fertility Decline." *Quarterly Journal of Economics* 89 (1, February): 1–31.

Lucas, Robert E. B., and Oded Stark. 1985. "Motivations to Remit: Evidence from Botswana." *Journal of Political Economy* 93 (5, October): 901–18.

Maddalla, G. S. 1984. *Limited-Dependent and Qualitative Variable in Econometrics*. Cambridge University Press.

McElroy, Marjorie B., and Mary Jean Horney. 1981. "Nash-bargained Household Generalization of the Theory of Demand." *International Economic Review* 22 (2, June): 333–49.

Mesa-Lago, Carmelo. 1985. *El Desarrollo de la Seguridad Social en America Latina*. Santiago: Comision Economica para America Latina y el Caribe.

Nugent, Jeffrey B. 1985. "The Old-Age Motive for Fertility." *Population and Development Review* 11 (1, March): 75–97.

Ravallion, Martin, and Lorraine Dearden. 1988. "Social Security in a 'Moral Economy': An Empirical Analysis for Java." *Review of Economics and Statistics* 70 (1, February): 36–44.

Rempel, Henry, and Richard A. Lobdell. 1978. "The Role of Urban-to-Rural Remittances in Rural Development." *The Journal of Development Studies* 14 (3, April): 324–41.

Rosenweig, Mark R. 1986. "Program Intervention, Intrahousehold Distribution, and the Welfare of Individuals: Modeling Household Behavior." *World Development* 14 (2, March): 233–43.

Suarez-Berenguela, Ruben M. 1987. *Financing the Health Care Sector in Peru*. LSMS Working Paper 31. Washington, D.C.: World Bank.

U.S. Social Security Administration. 1987. *Social Security Programs throughout the World*. Washington, D.C.: Government Printing Office.

Vlassoff, Michael, and Carol Vlassoff. 1980. "Old Age Security and the Utility of Children in Rural India." *Population Studies* 34 (3, November): 487–99.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 1: 171-18

The Willingness to Pay for Education for Daughters in Contrast to Sons: Evidence from Rural Peru

Paul Gertler and Paul Glewwe

In most of the developing world the education of women lags behind that of men. This could come about from a lack of parental desire for educated daughters or from a perception by the parents that there is a lower net return to education for girls. The relation between gender and education in rural Peru is explored using data from the 1985–86 Peru Living Standards Survey. A model of educational choice is developed. The estimated demand functions are used to assess the impact of user fees on demand and revenues. The empirical evidence indicates that parents are more willing to pay for reduced travel time to secondary school for boys than for girls. However, parents are willing to pay increased fees for girls' schooling sufficient to generate teachers' salaries.

Education is almost universally recognized as an avenue for raising living sta dards. Thus the factors that influence educational decisions have an importa impact on individual and social welfare. In developing countries one factor th plays a consistent role in educational decisions is gender: females usually atta lower levels of education and have lower school enrollment rates than do male In almost all developing countries the government is the main provider of educ tion, but in recent years funds for education have been constrained as mar developing countries have experienced economic stagnation. In times of tight educational opportunities the educational position of women, relative to me may suffer.

In many developing countries the government provides educational oppo tunities at no direct cost to households, or at very little cost relative to t expenses it incurs. Present financial burdens on governments have led some suggest that households should pay substantial fees to provide funds for t government to expand both the quantity and the quality of schooling (Birds 1983; Thobani 1983; World Bank 1986; Jimenez 1987). Whether such a poli would discriminate against poorer households has been investigated in Gertl and Glewwe (1989), who find that such households are willing to pay sizab

Paul Gertler is at the RAND Corporation, Santa Monica, California, and Paul Glewwe is in th Population and Human Resources Department at the World Bank.

Table 1. *Primary and Secondary School Enrollment Rates for Boys and Girls in Various Income and Country Groups and in Peru, 1986*
(percentage of age group)

Group	Primary Boys	Primary Girls	Secondary Boys	Secondary Girls
Income group				
Low	113	92	42	27
Lower-middle	108	100	57	50
Upper-middle	107	101	71	67
High	103	102	91	93
Country group				
Sub-Saharan Africa	73	58	20	12
East Asia	131	117	50	39
South Asia	98	69	41	22
Latin America and the Caribbean	110	108	54	56
Developed economies	103	102	92	94
Peru	125	120	68	61

Note: All rates are gross rates and thus may be greater than 100 because some of the children enrolled may be older than the age group normally associated with a given level of schooling.
Source: World Bank (1989).

fees to send their children to school. Yet there is little rigorous investigation of the effect of school fees on the educational attainment of girls relative to that of boys. This question is examined here using data from rural Peru. We find that although parents are less willing to pay for the schooling of girls relative to that of boys, they are still willing to pay a fee that would be adequate to finance teachers' salaries.

I. Low Levels of Education among Women—Theory and Evidence

Primary and secondary school enrollment rates for girls and boys in various income and country groups and in Peru are given in table 1. The gap between girls' and boys' enrollment rates is largest for the poorest countries; it effectively disappears for high-income countries, at least at the primary and secondary levels. The gap is most pronounced in Sub-Saharan Africa and South Asia, less pronounced in East Asia, and almost nonexistent in Latin America and the Caribbean and in developed economies. However, even in Latin America and the Caribbean, progress in school enrollment for girls is uneven across countries.

There are two main aspects of a decision by parents to send their children to school. The first is the consumption aspect, which is that parents may prefer educated children regardless of the financial benefits of education. The second is the investment aspect, which proposes that education may be valued because it brings financial (or other) returns to parents. Thus differences in school enrollment for girls and boys must occur because of differences in the direct value parents place on educated children or because the economy and society may be

such that, from the parents' perspective, the net return to educating boys is higher than that to educating girls. Of course, both factors may be operating.

Differences in the net returns to education can be divided into the cost of, and the gross returns to, education and into the returns to parents and the returns to children. The net return to education is the gross return minus the cost (both in terms of discounted value). If the net gains are lower for girls, either their costs are higher or their gross financial gains are lower relative to boys, or both. Furthermore parents may not benefit equally from financially successful sons and daughters. If social customs place more responsibility on sons than on daughters to support their parents, the parents may have a stronger incentive to educate their sons, even if the net returns to education are the same for boys and girls.

II. Gender and Education in Rural Peru

The relation between gender and education in rural Peru is explored using the Peru Living Standards Survey. This survey was carried out jointly by Peru's Instituto Nacional de Estadistica and the World Bank (see Grootaert and Arriagada 1986).

The Extent of the Gender Gap

Tables 2 and 3 present data on the differences in education by gender in urban and rural areas in Peru. The data on the adult population in table 2 indicate that in recent years the gap between women and men in the level of education

Table 2. *Mean Years of Schooling Attained and the Gender Gap in Education in Peru, 1985–86*

	Mean years of education			*Gender gap*	
Age	*All*	*Male*	*Female*	*Years*[a]	*Percentage*[b]
Adult population					
20–29	—	8.6	7.4	1.2	16.2
30–39	—	7.7	5.8	1.9	32.8
40–49	—	5.6	3.7	1.9	51.4
50–59	—	4.9	2.9	2.0	69.0
60+	—	4.0	2.4	1.6	66.7
School-age population in urban areas					
6–10	1.3	1.4	1.3	0.1	7.7
11–15	5.2	5.2	5.0	0.2	4.0
16–20	8.5	8.6	8.5	0.1	1.2
21–25	9.9	10.1	9.6	0.5	5.2
School-age population in rural areas					
6–10	0.8	0.8	0.8	0.0	0.0
11–15	3.5	3.6	3.3	0.3	9.1
16–20	5.5	5.7	5.1	0.6	11.8
21–25	5.8	6.1	5.5	0.6	10.9

— Not available.
a. Difference between male and female mean years of schooling.
b. The difference as a percentage of female mean years of schooling.
Source: Authors' calculations from the unpublished 1985–86 Peru Living Standards Survey.

Table 3. *Boys and Girls in the Primary and Secondary Grades in Urban and Rural Peru, 1985–86*
(percentage of students)

Grade	Urban		Rural	
	Boys	Girls	Boys	Girls
1	55.8	44.2	57.1	42.9
2	60.4	39.6	59.7	40.3
3	63.4	36.6	57.7	42.3
4	68.4	31.6	62.9	37.1
5	58.4	41.6	66.7	33.3
6	60.7	39.3	67.0	33.0
7	61.6	38.4	72.7	27.3
8	64.0	36.0	68.2	31.8
9	60.2	39.8	69.1	30.9
10	68.0	32.0	60.3	39.7
Primary (1–5)	60.6	39.4	59.9	40.1
Secondary (6–10)	62.5	37.5	67.7	32.3

Source: Authors' calculations from the unpublished 1985–86 Peru Living Standards Survey.

attained has narrowed. For the population age 40 and above, the mean years of schooling of males is more than 50 percent higher than that of females, but it is only 33 and 16 percent higher for males ages 30–39 and 20–29, respectively. For the school-age population in urban and rural areas, the data indicate that males and females have attained lower average years of schooling in rural areas. The gender gap is most obvious at the secondary level in rural areas.

Data on the percentage of boys and girls in grades 1 to 10 in urban and rural areas are presented in table 3. In both urban and rural areas in all grades more than half the students are boys. Girls make up the smallest percentage of students in rural areas in the secondary grades.

Cultural Attitudes and the Cost of Educating Girls

Do parents in rural areas of Peru think that education is less important for girls than for boys? Some observers have found evidence to support this hypothesis. According to Fernandez (1986, p. 3), in rural areas of Peru "families maintain preferential attitudes toward boys and discriminatory ones toward girls." More than 90 percent of rural women in Peru do attend school at some point in their lives but, at the secondary level, boys appear to go farther (table 2). This is consistent with the notion that girls need only a basic education to perform their social roles, whereas boys require higher levels of education. According to Vargas (1987, p. 3), "Many rural and urban families of the lower and middle classes prefer to send their sons to school and give their daughters a bare minimum of education and then put them in charge of domestic tasks."

One possible reason that boys go farther in school than girls is that the cost of sending girls to school may be higher. The direct cost of school fees is likely to be the same, but a more important cost may be hours of work lost to the household due to a child's school attendance. Fernandez (1986, p. 5) finds that "in a study

of peasants and seamstresses, 42 percent of 684 peasant women interviewed declared that they had left school because of their responsibilities on the farm and around the house." Data from the 1985–86 Peru Living Standards Survey for hours spent on schoolwork and on other work for children ages 10–14 and 15–18 are given in table 4. The data are for two areas of rural Peru, for students and nonstudents, and for girls and boys. Girls ages 15–18 in rural Costa and ages 10–18 in rural Sierra and Selva spend substantially more hours on work that is not related to school than do boys. More important, the loss in hours worked for a girl who attends school is higher than the loss for a boy who attends school.

Differences in the Benefits of Education for Males and Females

Given this cost disparity by gender, could there also be a benefit disparity? Stein (1972) presents anecdotal evidence that Peruvian households consider males to be "worth more" than females. The issue of whether the economic environment makes girls "less valuable" than boys is explored in this section by using studies based on data from the 1985–86 Peru Living Standards Survey.

Wages and education. One way in which educating boys could be "worth more" than educating girls is that the wages of men may be higher than those of women. More specifically, the increase in wages brought about by increases in schooling may be higher for boys than for girls. Stelcner, Arriagada, and Moock (1987) and King (1989) have examined the returns to schooling for men and women, respectively, in both urban and rural Peru. They estimate the percentage

Table 4. *Time Spent on Schoolwork and Other Work by Boys and Girls in Rural Peru, 1985–86*
(hours per week)

			Other work					
	Schoolwork		Outside the house		Housework		Total	
Area	Boys	Girls	Boys	Girls	Boys	Girls	Boys	Girls
Rural Costa								
Students								
Age 10–14	23.6	22.8	7.7	4.7	9.3	11.8	17.0	16.5
Age 15–18	20.7	17.8	12.7	12.0	7.8	10.2	20.5	22.2
Nonstudents								
Age 10–14	0.0	0.0	21.3	11.8	6.8	15.8	28.1	27.6
Age 15–18	0.0	0.0	26.8	12.3	5.0	26.5	31.8	38.8
Rural Sierra and Selva								
Students								
Age 10–14	24.4	25.0	8.7	8.8	9.7	12.8	18.4	21.6
Age 15–18	24.6	24.6	13.1	9.3	8.1	15.7	21.2	25.0
Nonstudents								
Age 10–14	0.0	0.0	25.1	25.7	8.7	17.9	33.8	43.6
Age 15–18	0.0	0.0	36.3	27.7	6.9	22.9	43.2	50.6

Source: Authors' calculations from the unpublished 1985–86 Peru Living Standards Survey.

[illegible] schooling. To the exten[illegible] women's wages are lower than those of men for reasons other than scho[illegible] similar percentage increases for both wages amount to a lower increase in [illegible] to women, relative to men, for an additional year of schooling.

Two results stand out from the research. First, at the primary level the re[illegible] to education to men and women are quite similar in both urban and rural a[illegible] and all estimates are statistically different from zero. Second, although th[illegible] turns to secondary education are always statistically different from zero for m[illegible] this is not the case for women. More strikingly, the point estimate of the ret[illegible] to secondary education for women in rural areas is very low (and insignifica[illegible]) compared with that for men.

These results suggest three things. First, there is a similar financial incentive t[illegible] send both girls and boys to primary school in both urban and rural Peru, so one would not expect discrimination against girls at the primary level in either area due to different (gross) returns to education. Second, although the point estimates are not precise for females, the rates of return to secondary school in urban Peru do not appear to vary by gender, so there is little incentive to discriminate against girls in urban areas. Third, there is an incentive to discriminate against girls and in favor of boys at the secondary level in rural areas, since rates of return to girls are not significantly different from zero while they are so for boys. Thus, even if rural parents had no previous bias against educating girls, the increase in wages for males and the lack of increase in wages for females as a result of increased schooling may persuade them to send boys, but not girls, to secondary school.

Returns to education in nonagricultural self-employment. Many men and women are not wage earners, but are self-employed, either in agriculture or nonagricultural enterprises. Moock, Musgrove, and Stelcner (1989) estimated the rates of return to schooling on the earnings of self-employed individuals working in nonagricultural family enterprises. Separate estimates were made for all enterprises, those run only by females, and those run at least in part by males, for Lima, other urban areas, and rural areas.

In urban areas estimated rates of return to education are significantly different from zero for most levels of education. However, the returns to secondary education in female-only firms in urban areas are not significantly different from zero. In rural areas there are no significant returns to anyone at any level of education. Thus there does not appear to be any financial gain to secondary education for self-employed women in either urban or rural areas.

Returns to education in agriculture. Another possibility is that education may be a profitable investment if it increases agricultural productivity. Jacoby (1989) takes up this issue in a study that estimates the returns to education on gross farm revenues in rural Sierra. Jacoby's results do not distinguish between males and females, but in general indicate that education has a substantial effect on

gricultural productivity only at the postsecondary level. Thus it is not surprising that rural women are less likely to have attended secondary school than rural men, who at least have significant returns to education for wage work.

III. The Value of Secondary Education for Males and Females in Rural Peru

From the data in section II, two points became clear about differences in education for boys and girls in rural Peru. The first is that girls appear less likely to attend secondary school than do boys. The second is that the returns to secondary schooling are almost nil for women, while for men secondary schooling has significant returns in the wage labor market. Thus secondary education may not be a rational investment for girls, but it does seem to be a rational investment for boys. Consequently, girls may be less likely to be enrolled than boys. However, education also has a consumption aspect, and the fact that many girls are enrolled in secondary school indicates that, for many families, secondary schooling for girls is valued. (It is also possible that it makes them more productive in work at home.)

A model of the demand for schooling is used to calculate the willingness of parents to pay for improvements in their children's schooling. The model is estimated to get a general picture of the differential value placed on boys' and girls' schooling and to examine what will happen if user fees are raised in rural Peru.

The Model

Because education is both a consumption good and an investment good, it is valued for its own sake and because it provides financial returns. The fact that some children are not attending school indicates that, in their parents' eyes, the advantages of sending them to school are outweighed by the disadvantages. The advantages are the benefits that parents receive from well-educated children, both in direct material returns (for example, support in old age) and in the satisfaction of having educated children. The main disadvantages are the associated costs, which may be a heavier burden on poorer families.

Parents plan for the total investment in a child's education based on assumptions about future costs and benefits. Because schooling occurs in annual increments, parents must decide each year whether to enroll their child in another year of school. As time passes and more information about expected costs and benefits is revealed, parents update their expectations and revise their plans.

Although it might be optimal to model the complete dynamic process, the limitations of a single cross-section of data require a focus on the decision to enroll for another year of schooling. The contemporaneous enrollment decision is chosen over the educational achievement outcome because we are interested in price effects and because the latter requires data on all past prices and incomes

when choices about education were made. However, the enrollment decision model is consistent with a fully dynamic process. Ultimately, a family enrolls a child for another year of school if the expected future benefits outweigh the associated costs.

A model of the demand for schooling is formalized by specifying the utility obtained from each schooling option (Gertler and Glewwe 1989). Every household is assumed to have a utility function that depends on the human capital of its children and the consumption of goods and services. Investment in another year of schooling raises a child's human capital at the cost of reduced consumption of other goods and services. In any year the choice is discrete; the options are to not send the child to school, to send the child to the nearest school, or to send the child away from home to a distant school.

The expected utility conditional on sending a child to school is given by

$$U_i = U[(H_0 + S_i), C_i] + \epsilon_i, \qquad i = l, f \tag{1}$$

where H_0 is the current stock of the child's human capital, S_i is the increment to that human capital from another year of education at school i, C_i is the household's consumption of goods and services after incurring the direct and indirect costs of sending the child to school i, and ϵ_i is a random taste shifter. Subscripts l and f refer to local and faraway (that is, the child lives away from home) schools, respectively. If the parents decide not to send the child to school, the utility is:

$$U_0 = U(H_0, C_0) + \epsilon_0 \tag{2}$$

where C_0 is the level of household consumption possible without sending the child to school, and ϵ_0 is a random taste shifter.

The decision to send a child to school depends on the quality of education received (Behrman and Birdsall 1983, 1985), which is determined by school and teacher characteristics, such as the availability of textbooks and levels of teacher training. The quality of education can also include household characteristics such as parents' education. It is assumed that the higher the quality of education, the greater the increase in human capital from another year of schooling and thus the higher the utility from the schooling option, other things being equal. One would expect that the only reason for a child to go to a faraway school would be to obtain a higher quality of education than is available at the local school.

The local and faraway school options have budget constraints of the form

$$C_i + P_i^* = C_0 = Y, \qquad i = l, f \tag{3}$$

where P_l^* and P_f^* are the prices of sending a child to local and faraway schools for one year, respectively, and Y is the family's yearly disposable income. The price of schooling involves both direct and indirect costs. The former include expenses such as school fees and outlays for textbooks. The crucial indirect cost is the opportunity cost of a child's time. In developing countries children often contribute substantially to family income by working, including housework. Table 4

demonstrates that there are large differences in hours worked between students and nonstudents in rural Peru.

Direct and indirect costs are not necessarily the same for all households. Even if fees and textbook costs are the same for all households, direct transportation costs may vary. If the school is far away, outlays on transportation may be required, and, if the school attended is so far that the child must live away from home, lodging costs are also incurred. Indirect costs may also vary because of differences in time lost from work due to school attendance and variations in wages earned by children. The variation in direct and indirect costs allows estimation of the effect of raising user fees on school enrollments.

Given the three schooling options and their associated utilities, the household's unconditional utility maximization problem is

$$U^* = \max(U_0, U_l, U_f) \tag{4}$$

where U^* is maximum utility, and U_0, U_l, and U_f are the conditional utility functions specified in equations 1 and 2, given the constraints in equation 3. The stochastic terms in the utility function imply that the solution to the utility maximization problem gives the probability that each alternative is chosen. In a discrete choice framework these probabilities can be interpreted as demand functions. These demand functions are used to solve for the unconditional indirect utility and expenditure functions and to assess the welfare impact of policy changes in terms of compensating variations.

Solving the utility maximization problem in equation 4 yields a system of demand functions reflecting the probabilities associated with each alternative. The functional form of each demand function depends on that of the conditional utility function and on the distribution of the stochastic variables. A parsimonious form of the conditional utility function that does not impose a constant marginal rate of substitution (and thus allows estimates of willingness to pay to vary by household income level) is the semiquadratic, which is linear in human capital and quadratic in consumption. The conditional utility function for the two schooling options is

$$U_i = \alpha_0(H_0 + S_i) + \alpha_1 C_i + \alpha_2 C_i^2 + \epsilon_i, \qquad i = l, f. \tag{5}$$

Each ϵ_i has zero mean and finite variance and is uncorrelated across individuals.

Consumption net of expenditures for schooling can be derived from equation 3 to yield:

$$C_i = Y - P_i^* = Y - P_i - wT_i, \qquad i = l, f \tag{6}$$

where w is the opportunity cost of the child's time; T_i is the child's lost hours of work due to school attendance, which varies over the local and faraway school options; and P_i is the direct price (money cost) of sending the child to school i. Substituting equation 6 into equation 5 yields

$$\begin{aligned} U_i = {} & \alpha_0(H_0 + S_i) + \alpha_1(Y - P_i - wT_i) \\ & + \alpha_2(Y - P_i - wT_i)^2 + \epsilon_i, \qquad i = l, f \end{aligned} \tag{7}$$

is the utility the household derives from [illegible] the child to either a loca[illegible] faraway school. The utility from keeping the child at home is

$$U_0 = \alpha_0 H_0 + \alpha_1 Y + \alpha_2 Y^2 + \epsilon_0. \tag{8}$$

The identification of the parameters in equations 7 and 8 requires that t[illegible] values of human capital and consumption differ across the alternatives. Vari[illegible] tion in prices for the two schooling options identifies α_1 and α_2 in equation [illegible] Because the decision rule involves comparing utility levels across alternatives, the conditional utility functions can be normalized relative to one of the alternatives without loss of generality. The utility from the no-schooling alternative is normalized to zero by subtracting equation 8 from equation 7 for both the local and faraway options. In this case equation 7 becomes

$$\begin{aligned} U_i - U_0 = {} & \alpha_0 S_i - \alpha_1 (P_i + wT_i) \\ & - \alpha_2 [2(P_i + wT_i) Y - (P_i + wT_i)^2] + \epsilon_i - \epsilon_0. \end{aligned} \tag{9}$$

The current stock of human capital is differenced out of the model. Income is differenced out of the consumption term, but it is not differenced out of the consumption-squared term. The linear consumption term represents just price, whereas the consumption-squared term includes both a price-income interaction term and a squared price term. If a linear utility function had been used, the squared terms in equations 7 and 8 would not be present, and the marginal rate of substitution across schooling alternatives would be constant. Income would then difference out of the decision rule, and consequently it would not influence the alternative chosen. In this specification the quadratic consumption term implicitly allows price effects to vary by income.

Not all schools are the same; better schools will provide "more" human capital per school year and thus be more attractive to parents. Quality cannot be observed directly, but one can observe various school and demographic characteristics thought to contribute to school quality and estimate a household human capital production function. However, the marginal utility of increments to human capital may also vary by demographic characteristics. Pollak and Wachter (1975) argue that the separate effects of demographic variables, such as parents' education, in the production of human capital and in the marginal utility function cannot be identified. Therefore a reduced-form model of the utility is specified from human capital accumulated by attending school i:

$$\alpha_0 S_i = \gamma X_i + \eta_i \tag{10}$$

where X_i is a vector of school quality and demographic characteristics. The coefficients in equation 10 differ for the local and faraway school alternatives. The random disturbance term, η_i, captures unmeasured portions of the quality function and may be correlated across alternatives.

School quality variables are components of the human capital production function, while demographic characteristics may enter both in the production function and as direct determinants of the marginal utility received from incre-

ents to human capital. Elements in the vector X_i that are specific to households or individuals will not vary across schools and thus do not take the i subscript. Those which are school-specific will vary by schools, hence the need for the i subscript on X.

Substituting equation 10 into the conditional utility function 7 yields

$$U_i = V_i + \epsilon_i + \eta_i, \qquad i = l, f \tag{11}$$

where $V_i = \alpha_0 H_0 + \gamma X_i + \alpha_1(Y - P_i - wH_i) + \alpha_2(Y - P_i - wH_i)^2$, and

$$U_0 = V_0 + \epsilon_0 \tag{12}$$

where $V_0 = \alpha_1 Y + \alpha_2 Y^2$ for the no-schooling option. V simply represents the nonstochastic component of the utility obtained from each option.

The Demand Functions and Welfare

The demand functions for schooling are assumed to take a nested multinomial logit form (McFadden 1981) to allow for correlation across subgroups of alternatives and, therefore, nonconstant cross-price elasticities. Specifically, the joint distribution of the error terms is assumed to follow a type B extreme value distribution. The error terms of the schooling alternatives (in equation 11) may be correlated with each other, but not with those in the no-school alternative. Thus the probability of not going to school is

$$\pi_0 = \frac{\exp(V_0)}{\exp(V_0) + [\exp(V_l/\sigma) + \exp(V_f/\sigma)]^\sigma}. \tag{13}$$

Intuitively, as the nonstochastic component (V_0) of the utility from the no-school option rises, the probability of not attending school increases. Similarly, as the nonstochastic component of the utility from either of the two schooling options (V_l and V_f) rises, the probability of not attending school declines. The demand for schooling alternative i is

$$\pi_i = (1 - \pi_0) \frac{\exp(V_i/\sigma)}{[\exp(V_l/\sigma) + \exp(V_f/\sigma)]^\sigma}, \qquad i = l, f \tag{14}$$

where σ is one minus the correlation of the error terms in the conditional utility functions of the local and faraway schooling options introduced by the η_i's, and the V_i's are given by equations 10 and 11. The intuition here is that once the probability of not going to school has been determined, the choice between the two schooling options is made. If the nonstochastic component of the utility from the local schooling option rises relative to the faraway schooling option, the probability of going to the local school increases through the increase in the fraction term in equation 14. An analogous situation holds if the nonstochastic component of the utility from the faraway schooling option increases.

The estimated demand functions in equations 13 and 14 can be used to assess the impact of user fees on demand and revenue and, consequently, to assess the willingness to pay for closer or higher-quality schools. Willingness to pay is

:alculated as a compensating variation—the amount of income needed to make he family just as well off after a price or quality change as before the change. In he case of a nested multinomial logit (Small and Rosen 1981), the compensat-ıg variation, denoted by CV, for an increase in travel time is

$$CV = (1/\lambda)\{ln\,[\exp(V_o) + [\exp(V_l/\sigma) + \exp(V_f/\sigma)]^{\sigma}] - ln\,[\exp(V_0') + [\exp(V_l'/\sigma) + \exp(V_f'/\sigma)]^{\sigma}]\} \tag{5}$$

'here V_i and V_i' are evaluated at the original and new travel times, respectively, ıd λ is the marginal utility of income. Intuitively, the change in the nonstochas-c component of utility is given by the difference between the two main terms in ıe curled brackets. The marginal utility of income is then used to obtain the ıoney value (compensating variation) of this change.

IV. Estimation of the Model

The above model of the demand for schooling was implemented using data om the Peru Living Standards Survey. The data set and construction of the ıriables are the same as in Gertler and Glewwe (1989).

The Variables

"Consumption" is the value of per capita consumption associated with each ›tion. It equals family income under the no-schooling option, and family in-ıme minus the direct and indirect costs of schooling under the other options. onsumption is observed under the current choice. Thus if the child is in school, mily income is calculated as the sum of observed consumption and the price of e schooling option chosen. The indirect cost of the schooling option is the 'oduct of the local wage rate for children and the difference in hours worked by udents and nonstudents, as given in table 4. If a child goes to a faraway school, is assumed that all the child's labor is lost.

The direct (monetary) cost of schooling was calculated from regional means expenditures on tuition and books, plus mean transportation costs calculated r three different categories of travel time. Estimates were also done using 'edicted hours worked by students and nonstudents in order to account for mple selectivity that might be present in the figures given in table 4. Evidence r selectivity was weak, and the results were essentially unchanged (Gertler and lewwe 1989).

The age of the child is included as a variable because older children may be ss inclined to continue schooling at a given grade level. "Year in school" is the ımber of school years completed by the child. It is included to see if some ades of secondary school are perceived to be more valuable than others. A ımmy variable is included that takes the value of 1 if the child is female and 0 if e child is male.

The number of years of completed schooling by the child's father and mother e also included. Parental education measures preferences for education, possi-

ble ability of children, and assistance given by parents to children in schoolwork. Another family-characteristic variable that is included is the number of siblings between the ages of 13 and 17.

"Travel time" is interpreted as a school quality variable. Longer travel times to local schools reduce the hours of time spent by children in school and thus reduce the quality of education they receive. The other school quality variable is a lack of teachers at the local primary school, which is assumed to be highly correlated with a lack of teachers in local secondary schools. Unfortunately, there are no other variables that indicate quality of education or school quality.

The Estimates

Estimates of the model of educational choice for secondary school in rural Peru are given in table 5. The consumption and consumption-squared coefficients indicate that the demand for schooling rises with expenditure levels but is nonlinear. The estimate of the correlation across the error terms is between 0 and 1, which indicates that the nested multinomial logit model is correctly specified. The other parameter estimates were allowed to vary by choice of schooling but in some cases have similar magnitudes. Age has a significantly

Table 5. *Estimation Results for the Nested Multinomial Logit Educational Choice Model*

Independent variable	*Coefficient*	*t-statistic*
Consumption[a]	0.5936	2.0621
Consumpton squared[b]	−0.0250	−1.6554
Correlation term (sigma)	0.6865	3.3285
Local schooling option		
Constant	3.5229	2.3672
Age	−0.5115	−3.0834
Year in school	0.6121	2.3225
Female	−0.9542	−2.5174
Father's education	0.0375	0.6045
Mother's education	0.1595	1.7885
Siblings age 13–17	0.0141	0.0829
Travel time	−0.4270	−5.4625
Teacher shortage	−1.3419	−2.7956
Faraway schooling option		
Constant	1.6774	1.0167
Age	−0.4561	−2.6447
Year in school	1.0458	3.5927
Female	−0.9472	−2.2289
Father's education	−0.0006	−0.0088
Mother's education	0.0259	0.2729
Siblings age 13–17	−1.0731	−4.6619
Number of observations	718	
− Log likelihood	569.65	

a. Variable divided by 100 for estimation.
b. Variable divided by 10,000 for estimation.
Source: Authors' calculations from the unpublished 1985–86 Peru Living Standards Survey.

negative effect on the demand for schooling in both options, indicating that conditional on finishing primary school, the demand for schooling is stronger for younger children. There is clearly a lower demand by parents for both types of schooling for female children. The father's education has little effect on the demand for either type of schooling. Finally, "year in school" has a strong positive effect on the demand for both types of schooling, which suggests that the final years of secondary school are more valuable than the initial years, all things being equal.

The other parameter estimates varied by schooling options. The mother's education has a sizable positive effect on the demand for the local schooling option but has almost no effect on that for the faraway schooling option. The presence of other siblings of the same age, which is a proxy for other demands on scarce household resources, has no effect on the demand for the local schooling option but has a strongly negative effect on the faraway option. Finally, the two variables indicating quality of education—travel time (which implies less classroom time) and a dummy variable (which indicates lack of teachers in the local primary school)—are significantly negative.

The parameter estimates in table 5 are used in table 6 to predict the probability of choosing the three schooling options by gender and by household expenditure level per capita. All other variables are evaluated at their mean levels. At all expenditure levels the probability of males being in local schools is higher than that for females, and the same is true for probabilities of being in faraway schools. For example, at the level of 5,000 intis per capita per month, the probability of a girl attending a local school is about 0.19, whereas for a boy it is 0.28, and the probabilities for attending a faraway school are 0.13 and 0.19 for girls and boys, respectively. For both girls and boys the probability of going

Table 6. *Probability of Choosing Various Schooling Options for Males and Females by Household Expenditure Level*

Household expenditure level (intis per capita per month)	*Males*			*Females*		
	Local school	*Far-away school*	*No school*	*Local school*	*Far-away school*	*No school*
1,000	0.280	0.166	0.554	0.185	0.110	0.705
2,000	0.281	0.173	0.547	0.186	0.115	0.699
3,000	0.281	0.179	0.540	0.187	0.120	0.693
4,000	0.281	0.186	0.533	0.188	0.125	0.687
5,000	0.281	0.193	0.526	0.189	0.131	0.681
6,000	0.281	0.200	0.519	0.190	0.136	0.674
7,000	0.281	0.207	0.511	0.191	0.142	0.668
8,000	0.281	0.215	0.504	0.191	0.147	0.661
9,000	0.281	0.223	0.497	0.192	0.153	0.655
0,000	0.280	0.230	0.489	0.193	0.159	0.648

Source: Authors' calculations from the unpublished 1985–86 Peru Living Standards Survey.

Table 7. *Arc Price Elasticities by Gender*

Range of increase in school fees (intis)	Quartiles of household income level: Lowest	Next lowest	Next highest	Highest	All households
Average household income	7,236	13,154	21,002	48,679	20,483
Males					
Local school					
0–300	−0.19	−0.18	−0.16	−0.09	−0.18
300–600	−0.35	−0.33	−0.29	−0.16	−0.32
600–900	−0.53	−0.49	−0.44	−0.24	−0.47
900–1,200	−0.73	−0.67	−0.59	−0.33	−0.64
Faraway school					
0–300	−0.22	−0.21	−0.19	−0.11	−0.18
300–600	−0.40	−0.37	−0.33	−0.19	−0.33
600–900	−0.59	−0.55	−0.49	−0.27	−0.49
900–1,200	−0.80	−0.74	−0.66	−0.36	−0.66
Females					
Local school					
0–300	−0.22	−0.20	−0.18	−0.11	−0.19
300–600	−0.39	−0.36	−0.32	−0.19	−0.32
600–900	−0.57	−0.53	−0.47	−0.27	−0.48
900–1,200	−0.78	−0.72	−0.64	−0.36	−0.65
Faraway school					
0–300	−0.24	−0.22	−0.20	−0.12	−0.20
300–600	−0.43	−0.39	−0.35	−0.20	−0.36
600–900	−0.63	−0.58	−0.51	−0.29	−0.52
900–1,200	−0.85	−0.78	−0.69	−0.38	−0.70

Note: Arc elasticities for a given price range are defined as the percentage change in demand (in this case the probability of attending a certain school) divided by the percentage change in price. The two endpoints of the price range are used to calculate the percentage change in the price.

Source: Authors' calculations from the unpublished 1985–86 Peru Living Standards Survey.

to school rises with expenditure levels. For girls this is true of both local schools and faraway schools; for boys the probability of attending a local school is essentially unchanged, whereas that of attending a faraway school rises substantially. Although households that are better-off are more likely to send both girls and boys to school, the gap between girls and boys is found among all households and does not disappear for the better-off rural households in Peru.

The continued preference for sending boys to school also appears in the estimates of arc price elasticities of the demand for schooling given in table 7. For both boys and girls, arc price elasticities are less than unity in absolute value, indicating that the demand for schooling is relatively insensitive to price changes. The elasticities are lower for wealthier households but increase at higher levels of proposed increases in school fees, for both girls and boys. Price elasticities for girls are always higher in terms of absolute value, which indicates that the demand for schooling for girls is more sensitive to price increases than for boys. This suggests that increases in user fees may reduce schooling more for girls than for boys. But Gertler and Glewwe (1989) argue that price elasticities alone do not indicate whether parents are willing to pay to improve school

Table 8. Annual School Fee Parents Would be Willing to Pay to Reduce Travel Time by Two Hours in Rural Peru
(June 1985 intis)

Household expenditures per capita (intis per month)	*Male*	*Female*
1,000	440	374
2,000	486	414
3,000	542	462
4,000	612	523
5,000	702	602
6,000	823	702
7,000	994	855
8,000	1,253	1,080
9,000	1,691	1,462
10,000	2,600	2,252

Source: Authors' calculations from the unpublished 1985–86 Peru Living Standards Survey.

quality or to construct new schools closer to their communities. Similarly, they do not necessarily indicate that parents are less willing to pay for improvements in schooling for girls than they are for boys.

Estimates of the willingness to pay for improvements in schooling are given in table 8, which shows how much parents are willing to pay to reduce travel time to the nearest secondary school by two hours. If the government were to build a new school that reduced travel by this amount, the full cost of supporting teacher salaries would require an annual school fee of 400 intis (Gertler and Glewwe 1989). With one borderline exception, parents of both boys and girls are willing to pay this cost for closer schools.

Table 8 shows that parents are more willing to pay for reduced travel time for boys than for girls, and that this holds for both poor and better-off households. Yet the differences here are not astounding; parents are still quite willing to pay to improve schooling for girls even though there is evidence that the utility of secondary education for women is less than that for men in rural Peru. Averaging across household expenditure levels, parents are willing to pay only about 17 percent more for boys than for girls. This indicates that parents value secondary education for girls even though economic returns to it in the labor market are doubtful.

Several interpretations of these results, not necessarily mutually exclusive, are possible. First, secondary education for girls may be useful in their work around the home. Second, some parents may expect their daughters to migrate to urban areas, where there are substantial returns to secondary education for female wage earners. Third, to the extent that returns to higher education are significant for women (Moock, Musgrove, and Stelcner 1989; Jacoby 1989), some parents may expect their daughters to continue their education beyond secondary school and thus are willing to pay for secondary education. Fourth, educated daughters may simply be directly valued by many parents.

V. Conclusion

For developing countries whose budgetary resources are limited, raising revenue for education by raising school fees is a policy option. There is much anecdotal evidence that girls are not treated favorably, and it appears that the returns to secondary education for girls in rural Peru are not very strong. Examination of the empirical evidence indicates that parents in rural Peru place more value on sending boys to secondary school than on sending girls. However, there is adequate demand for girls' schooling so that raising user fees to generate teachers' salaries (for a nearer school) will not result in a welfare loss that will cause parents to remove their daughters from school. This is true even among relatively poor rural households.

Whether this is true in other countries remains to be seen; rigorous empirical work along the lines done here is needed. It is possible that larger gender differences in the willingness to pay for schooling may lead to a marked decrease in school attendance among girls. The obvious policy remedy is to set lower school fees for girls relative to boys. This type of price discrimination will allow for more flexibility in raising revenue to expand educational opportunities while at the same time maintaining educational opportunities for girls. Such a policy has recently been adopted in parts of Bangladesh and Nigeria and should be taken seriously in any developing country where providing educational opportunities to girls is a high priority.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Behrman, Jere, and Nancy Birdsall. 1983. "The Quality of Schooling." *American Economic Review* 73 (5): 928–46.

———. 1985. "The Quality of Schooling: Reply." *American Economic Review* 75 (5): 1202–05.

Birdsall, Nancy. 1983. "Strategies for Analyzing Effects of User Charges in the Social Sectors." Discussion Paper 1983-9. World Bank, Country Policy Department, Washington, D.C. Processed.

Fernandez, Hernan. 1986. "Women's Educational Situation in Peru." Instituto Nacional de Investigacion y Desarrollo de la Educación, Lima, Peru. Processed.

Gertler, Paul, and Paul Glewwe. 1989. *The Willingness to Pay for Education in Developing Countries: Evidence from Rural Peru.* LSMS Working Paper 54. Washington, D.C.: World Bank.

Grootaert, Christiaan, and Ana-Maria Arriagada. 1986. "The Peru Living Standards Survey: An Annotated Questionnaire." World Bank, Population and Human Resources Department, Washington, D.C. Processed.

Jacoby, Hanan. 1989. "The Returns to Education in the Agriculture of the Peruvian Sierra." World Bank, Washington, D.C. Processed.

Jimenez, Emmanuel. 1987. *Pricing Policies in the Social Sectors*. Baltimore, Md.: Johns Hopkins University Press.

King, Elizabeth. 1989. *Does Education Pay in the Labor Market? The Labor Force Participation, Occupation, and Earnings of Peruvian Women*. LSMS Working Paper 67. Washington, D.C.: World Bank.

McFadden, Daniel. 1981. "Econometric Models of Probabilistic Choice." In C. Manski and D. McFadden, eds., *Structural Analysis of Discrete Data with Econometric Applications*. Cambridge, Mass.: MIT Press.

Moock, Peter, Philip Musgrove, and Morton Stelcner. 1989. "Education and Earnings in Peru's Informal Nonfarm Family Enterprises." PRE Working Paper 236. World Bank, Population and Human Resources Department, Washington, D.C. Processed.

Pollak, R., and M. Wachter. 1975. "The Relevance of the Household Production Function and Its Implications for the Allocation of Time." *Journal of Political Economy* 83:255–99.

Small, Kenneth A., and Harvey S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrica* 49 (1): 105–30.

Stein, William. 1972. "Mothers and Sons in the Andes: Developmental Implications." State University of New York at Buffalo, Department of Anthropology. Processed.

Stelcner, Morton, Ana-Maria Arriagada, and Peter Moock. 1987. *Wage Determinants and School Attainment among Men in Peru*. LSMS Working Paper 38. Washington, D.C.: World Bank.

Thobani, Mateen. 1983. *Charging User Fees for Social Services: The Case of Education in Malawi*. World Bank Staff Working Paper 572. Washington, D.C.

Vargas, Virginia. 1987. "Reflecting on Women's Education in Peru." Centro Flora Tristan, Lima, Peru. Processed.

World Bank. 1986. *Financing Education in Developing Countries: An Exploration of Policy Options*. Washington, D.C.

———. 1989. *World Development Report 1989*. New York: Oxford University Press.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 2: 189-211

Linking Trade and Productivity: New Research Directions

James R. Tybout

It is a mistake to think of productivity growth as an orderly shift in the production function of the representative plant. Gradual processes of technological diffusion or the displacement of inefficient plants with efficient ones are what matter. Trade orientation may affect these processes through many channels. Exposure to increased foreign competition is found to be associated with improvements in the average level of technical efficiency, reductions in the cross-plant dispersion in technical efficiency, and reductions in plant size. However, preliminary work suggests no clear link between trade policies and patterns of entry and exit.

This article brings together diverse literatures on the measurement of productivity and its relation to trade regime, focusing on recently developed techniques and their application. Section I briefly reviews the theoretical arguments linking trade policy and productivity. Empirical work at the sector and macro level is discussed in section II, and plant level empirical work is discussed in section III. Applications of the different approaches are reported for a sample of semi-industrial countries that have recently been analyzed in the World Bank research project "Industrial Competition, Productivity, and Their Relation to Trade Regimes," hereafter the ICPT project.

I. The Theory of Productivity Growth and Its Link to Trade

In trade models that presume perfect competition, "opening up" generally improves the allocation of factors across sectors and thereby induces a one-time increase in the value of domestic production. However, liberalization does not reduce the volume of inputs needed to produce a given bundle of outputs. Thus, although many economists believe that there are important links between trade

James R. Tybout is with the Department of Economics, Georgetown University. This article was prepared under funding for the World Bank research project "Industrial Competition, Productive Efficiency, and Their Relation to Trade Regimes," RPO 674-46. The author has benefited from the comments of Bela Balassa, Ricardo Caballero, Ann Harrison, Jaime de Melo, Mark Roberts, and two anonymous referees.

regimes and factor productivity, they have had to look elsewhere for formal models that support their priors.

Linking Trade Regimes and Productivity Levels

To development economists, perhaps the best-known argument linking trade regimes and productivity is that the returns to entrepreneurial effort increase with exposure to foreign competition (Corden 1974; Martin and Page 1983).[1] *Unfortunately, formal representations of this argument reveal its fragility. To hold, the entrepreneurial labor supply curve must be upward sloping in the relevant range, and changes in work incentives must go in the same direction for both export-oriented and import-substituting producers (Corden 1974; Rodrik 1988).*

Arguments based on increasing returns are also common in the development literature. Nishimizu and Page (1991, p. 253) summarize the logic as it has often appeared: "The existence of economies of scale . . . implies that a widening of the market through trade should lead to reductions in real production costs. In the context of an output-oriented development strategy, this argument is usually cast in terms of the benefits of increased demand through export expansion . . . ." As with entrepreneurial effort arguments, however, analytical scrutiny has shown that scale economies can cut both ways (Krugman 1986; Rodrik 1992; Roberts and Tybout 1991). When domestic firms enjoy market power, extra competition from foreign producers can force producers to expand or exit. But the net effect of liberalization on productivity depends upon the specifics of the demand shifts that accompany liberalization, ease of entry or exit, and the nature of competition.

Finally, trade reforms affect the tightness of the link between domestic and world markets and generate speculation about their own sustainability. These uncertainty effects can influence productivity. For example, when the incentive structure changes frequently and unpredictably, managers are reluctant to repeatedly incur the sunk costs of retooling (Dixit 1989a, 1989b; Baldwin and Krugman 1989; Baldwin 1989). Similarly, when substitution possibilities exist, managers may react to uncertainty by choosing labor-intensive technologies, even though more capital-intensive technologies would be less costly to operate if market conditions were stable (Lambson 1989). Rapid and efficient adjustments in productive capacity are likely only when trade reforms establish a credible, stable regime.

Diffusion and Innovation in Partial Equilibrium

Trade regimes have also been directly linked to rates of diffusion and innovation. For example, Rodrik (1992) shows that a firm's market share can affect the payoff it reaps from adopting new technologies. Trade reforms may therefore

1. Hart (1983) uses a principal-agent model to arrive at a similar result. His arguments are more subtle than those in the development literature, but equally fragile, as Scharfstein (1988) shows.

reduce the rate of catch-up to international productivity levels in import-competing sectors and accelerate it among exportables. Rodrik (1992) also observes that domestic producers compete through choice of technique (among other things), so they may tacitly collude when protected from foreign competition by failing to modernize their plants. He is quick to note that both arguments hinge critically upon some fairly arbitrary assumptions.

A more venerable strand of the literature focuses on plant heterogeneity. It begins with the premise that new processes diffuse through an industry as managers learn about them and older vintage machines depreciate. This means that there is no single production function, and it is a mistake to think of productivity growth as an orderly shift in technology. Rather, the processes of learning, innovation, investment, entry, and exit are what matter.[2] Trade orientation affects these processes through many channels, often by influencing entrepreneurial ability to monitor new technological developments or by changing the expected returns from innovation. Stewart and Ghani (1992) provide a useful review of the conceptual and empirical studies relevant to developing countries.

The New Literature on Endogenous Growth

Other strands of the literature provide general equilibrium representations of the links among trade, innovation, and growth. Unlike the treatments reviewed above, these models rely critically on the interaction among sectors. Examples are Krugman (1985) and Lucas (1988), which use learning-by-doing externalities at the sector level to create a link between trade policy and sectoral growth patterns. Any policy-induced shift in the composition of output changes patterns of sectoral learning rates and productivity growth; in turn these changes determine trade patterns.

Grossman and Helpman (1989, 1990) and Helpman (1990) explore another, related mechanism. In their models productivity growth is driven by private sector research and development, which results in new intermediate goods that enhance final good productivity and also contributes to public knowledge. Entrepreneurs in the research and development sector sell blueprints for income, and the rate of increase in the stock of blueprints determines the rate of productivity growth.[3]

Trade policy matters in this framework for several reasons. First, when deciding whether to develop new products, entrepreneurs consider the variety of substitute products already available, which depends in turn upon their exposure to international competition and the ease with which knowledge crosses international boundaries. Second, in larger markets there is more demand for

2. Jovanovic and Lach (1989) develop a partial equilibrium model of this view of innovation and diffusion.

3. In Grossman and Helpman (1989) the industrial countries develop new products, and developing countries invest in "de-engineering" them. This variant of the basic structure allows for product cycle effects.

any particular new product variety, so, other things being equal, market size encourages innovation. These two effects can work against one another, making the net impact of integrating with world markets ambiguous. Finally, research and development requires labor and capital inputs, which are also used to produce traded goods. So a change in trade regime that affects relative output prices also affects the returns to new product development (through Stolper-Samuelson linkages) and thereby influences the rate of productivity growth.

II. Sectoral and Industry-Level Approaches to Measuring Productivity

As the discussion above makes clear, there are many potential links between trade and productivity. It is not at all obvious which ones are empirically relevant, much less what their net effect will be in a particular liberalization episode. Accordingly, considerable attention has focused on empirical research.

Traditional Residual-Based Calculations

The most common approach to productivity measurement begins by assuming a neoclassical production function at the sectoral or industry level:

$$Y = f(\mathbf{v}, t). \tag{1}$$

Here total output (Y) is a concave function of the vector of inputs ($\mathbf{v}_{k \times 1}$) and a time index (t) that allows the function to shift with technological innovations or improvements in the efficiency of existing technologies. The elasticity of output with respect to time, $\epsilon_{Y,t} = (\partial f / \partial t)/Y$, is hereafter referred to as total factor productivity (TFP) growth.

The role of TFP growth is typically isolated by expressing equation 1 in growth terms and rearranging:

$$\epsilon_{Y,t} = \dot{Y}/Y - \sum_{j=1}^{k} \theta_j (\dot{v}_j / v_j). \tag{2}$$

Here a dot over a variable denotes its total derivative with respect to time, and $\theta_j = (\partial f / \partial v_j)(v_j / Y)$ is the elasticity of output with respect to the jth factor input. Then, making the critical assumption that each factor is paid the value of its marginal product, one may replace output elasticities with factor shares (s_j) and estimate TFP growth using a Divisia index:

$$\hat{\epsilon}_{Y,t} = \dot{Y}/Y - \sum_{j=1}^{k} s_j (\dot{v}_j / v_j). \tag{3}$$

The carat on $\epsilon_{Y,t}$ indicates that this is an estimator, and implementation requires that instantaneous time derivatives be replaced with discrete changes. The shares become averages of current and previous period shares, and the resultant measure of TFP growth is known as a Tornqvist index. In more involved applica-

tions diverse types of labor, capital, and intermediates are aggregated using Tornqvist indexes, and changes in the quality of each factor are analyzed (for example, Jorgenson and others 1987).

Given the possible trade-productivity links reviewed in section I, it has often been asked whether estimated TFP figures ($\hat{\varepsilon}_{Y,t}$) correlate with exposure to foreign competition. Much of this literature is surveyed elsewhere (Chenery and others 1986; Pack 1988; Havyrlyshyn 1990), so a few summary remarks suffice here. First, many cross-country studies find that rapid output growth is associated with rapid export growth or high export to output ratios. It is less common to find that rapid TFP growth correlates positively with openness. Pack (1988) writes that "comparisons of total factor productivity growth among countries pursuing different international trade orientations do not reveal systematic differences in productivity growth in manufacturing, . . ." However, Chenery and others (1986), Balassa (1985), and Edwards (1989) have found a positive association between TFP growth and openness. Second, after reviewing studies based on within-country temporal correlations, Pack (1988) and Havyrlyshyn (1990) both conclude that there is no strong evidence linking productivity and openness. Nonetheless some studies do find a positive association between export growth and productivity (for example, Krueger and Tuncer 1982; Nishimizu and Robinson 1984; Nishimizu and Page 1991). Third, in their multicountry study of industry-level TFP indexes, Nishimizu and Page (1991) find that other dimensions of policy—notably the degree of government intervention—significantly influence the relation between trade and productivity.

As a first step in researching the trade-productivity link, each author preparing a country study for the ICPT project was asked to regress annual industry-specific estimated TFP series on components of a demand-side sources of growth decomposition (domestic market growth, import substitution, and export expansion), a Herfindahl index of concentration, and an interaction term that allows the coefficient on import penetration to vary with market concentration. The index of concentration and interaction term allow for market structure effects like those discussed in Rodrik (1992)—they were a novelty of the ICPT specification. Industry dummies and annual time dummies were included in all regressions. Thus all explanatory variables were at the sector level.

The strongest result was a familiar one: output expansion and TFP growth covary. This positive association between the components of output growth and measured productivity, known as Verdoorn's Law, is sometimes taken to reflect scale economies or the embodiment of new technologies during periods of rapid investment (Chenery and others 1986; Nishimizu and Page 1991). Alternatively it may simply reflect spurious correlation due to the identity $X_1 + X_2 + X_3 = \hat{\varepsilon}_{Y,t} + \Sigma s_j(\dot{v}_j/v_j)$, where X_1, X_2, and X_3 decompose output growth into domestic market growth, import substitution, and export expansion, respectively. That is, if total output is measured with error, this error will show up on both sides o

he regression equation. Finally, as will be discussed shortly, violations of the ıssumptions behind equation 3 can induce spurious correlation.

The ICPT results also indicated that it matters whether demand expands be- 'ause of domestic market growth, export growth, or import substitution, but he pattern was country specific. In Chile and Morocco import substitution had ı significantly smaller effect on productivity growth in concentrated industries. ⅼut in Colombia import substitution was associated with especially high produc- ivity growth in concentrated industries. Taken together, the results suggest that narket structure does affect the nature of the link between trade patterns and ›roductivity, but the relationship is not a stable one. Methodological problems ıo doubt account for some of the instability; for example, measurement error nay occur in different components of output growth in different countries. But t is difficult to explain why these problems should cause the pattern to vary so trikingly across countries, or how they might cause market concentration to ondition the relation between import substitution and growth.

A Refinement of Residual-Based Productivity Measures

Even when data are observed without error, for calculations based on equa- on 3 to reflect productivity growth, the following assumptions are necessary: ıere are constant returns to scale, all factors are freely adjusted to maximize rofits, markets are competitive, and all plants employ identical technologies. ⅼone of these assumptions is innocuous. An extended version of this article rovides details on problems that result when these assumptions are violated nd reviews various corrections that have been attempted (Tybout 1991). ⅼerein we limit discussion to recent studies that attempt to control for noncon- :ant returns to scale, adjustment costs, and noncompetitive product markets ⅼall 1986, 1988a, 1988b; Domowitz and others 1988; Shapiro 1987; Cabal- 'ro and Lyons 1989a, 1989b, 1990).

Suppose that all plants are identical, and output price at the representative lant is given by the downward-sloping demand function, $P = p(Y)$ (with plant ıbscripts suppressed). Then the first-order conditions for profit maximization ղuate factor prices to marginal revenue products:

$$w_j = P[1 + \mu]\partial f/\partial v_j, \qquad j = 1, \ldots k \tag{4}$$

'here $\mu < 0$ is the inverse of the elasticity of demand perceived by the represen- ıtive plant, and $1/[1 + \mu]$ is its markup of price over marginal cost. Given that is negative, factor shares understate the true marginal product of the associ- ;ed factors, so TFP calculations based on equation 2 yield:

$$\hat{\epsilon}_{Y,t} = \dot{Y}/Y - (1+\mu)\sum_{j=1}^{k}\theta_j(\dot{v}_j/v_j) = \epsilon_{Y,t} - \mu\sum_{j=1}^{k}\theta_j(\dot{v}_j/v_j). \tag{5}$$

ccordingly, the extent of the upward bias depends directly upon the rate of :owth in factor stocks, and measured productivity growth is procyclical.

Hall (1986, 1988a, 1988b) suggests that this problem can be corrected with-

out knowing μ by regressing output growth on the share-weighted sum of input growth rates. Defining v_1, v_2, and v_3 as labor, materials, and capital, respectively, this amounts to fitting the regression:

$$\dot{Y}/Y = \beta_0 + \beta_1 \sum_{j=1}^{3} s_j(\dot{v}_j/v_j) + u. \tag{6}$$

By equation 4, $s_1 = \theta_1(1 + \mu)$, so equation 6 is a restatement of equation 2 with $\beta_1 = 1/(1 + \mu)$ and $\beta_0 = \epsilon_{y,t} - u$. This means that β_0 is the average rate of productivity growth, and β_1 is the markup over marginal cost.

The Hall methodology is appealing, but for several reasons its validity hinges critically on the availability of good instrumental variables. First, although markups are treated as parametric in equation 4, they are likely to be procyclical (Domowitz and others 1986), so ordinary least squares (OLS) estimates of β_1 are likely to be biased upward. Similar biases will result whenever the transitory TFP growth (u) is correlated with factor stock growth. Hall recognizes the potential correlation of factor stocks with the disturbance term, and uses growth of gross national product as an instrument in a two-factor version of equation 6. But critics of the Hall methodology argue that the problem remains, given that this growth is itself likely to be correlated with u and ϵ (Abbott and others 1989).

Each country study author in the ICPT project was asked to estimate a three-factor version of equation 6, generalized to the case of quasifixed capital stocks and nonconstant returns to scale. The most straightforward variant used was

$$\dot{Y}/Y - \dot{v}_3/v_3 = \beta_0 + \beta_1 \sum_{j=1}^{2} s_j(\dot{v}_j/v_j - \dot{v}_3/_3) + \beta_2(\dot{v}_3/v_3) + u. \tag{7}$$

Here $\beta_2 = 1 - \Sigma\theta_j$ is positive (negative) if firms exhibit increasing (decreasing) returns to scale. Most authors fit a version of this equation twice—once with OLS and once using instruments. Also, for two of the countries (Turkey and Côte d'Ivoire) the authors investigated whether trade reforms were associated with a change in either the corrected TFP growth rate or the price-cost markup. They did this by allowing β_0 and β_1 to shift in sample years when the economy was under a liberal trade regime.

All country authors performed regressions pooling time series on each three-digit industry (industry dummies were included). These typically yielded believable TFP growth figures (β_0) between 7 and −7 percent. They also yielded sensible markup coefficients (β_1), on average, ranging from 1.03 to 1.08 for OLS and 1.24 to 1.67 for instrumental variable estimators. But returns to scale (β_2) were not accurately identified, and the markup coefficients were especially sensitive to the instrument set used.

In separate ICPT studies Harrison (1990) and Levinsohn (1991) applied Hall's methodology at the plant level, where they could control for a number of complicating factors. They both concluded that certain protected sectors had significant markups and that these markups fell with trade liberalization or exchange rate appreciation. Harrison also found that, although TFP series based on equa-

tion 3 correlate strongly with trade regimes, the corrected TFP series based on equation 7 (that is, $\hat{\beta}_0$) do not.

Overall the Hall methodology appears to hold some promise, although it has several potentially serious shortcomings. First, it attributes all deviations of factor shares from marginal productivities to product market distortions. Second, it is disturbingly sensitive to the choice of instruments. Finally, it fails to deal with several of the same problems that limit the usefulness of Tornqvist indexes. Particularly when applied at the industry level, it involves heroic assumptions about the uniformity of technologies and behavior across plants as well as the homogeneity of factor inputs, both across plants and through time.

III. Micro Approaches to Measuring Productivity

The empirical methodologies discussed above all presume that a well-defined production technology describes all plants within the industry, sector, or country of analysis. If technological innovation takes place through a gradual process of efficient plants displacing inefficient ones or through the diffusion of new knowledge, the approaches to measuring productivity based on the behavior of a representative plant are at best misleading. At worst they fail to capture what is important about productivity growth altogether (Nelson 1981).

Heterogeneity and Productivity Growth

One of the most obvious features of industrial censuses is the tremendous amount of cross-plant heterogeneity. Even within narrowly defined industries, one observes wide ranges of output levels, capital-labor ratios, capital stock vintage, and profitability (Berry 1992; Tybout and others 1991). Accordingly, if changes in productivity are systematically induced by changes in the cross-plant distribution of these features, the productivity growth process will not be revealed by sectoral or macro analysis. An important question, about which we know very little, is whether such changes in the distribution of plants account for a significant portion of observed changes in sectoral output per unit of input (see, however, Gomulka 1976).

A complete analysis of this issue would require extensive engineering studies and is well beyond the scope of this article.[4] To provide a modest start, some simple statistics can be constructed to summarize the roles of plant turnover, scale, and heterogeneity in shaping productivity growth. Specifically, for a given industrial sector let $F = h(\mathrm{v})$ be a scalar index of factor input use. (For example, F might simply be number of workers, or it might be a share-weighted aggregation over capital, labor, and intermediates.) Then total industrial output (Y) can

4. Pack (1992) surveys the literature on engineering studies of certain groups of plants. He finds that "the emphasis in the recent technical change studies on the firm rather than the industry makes it difficult to evaluate the significance of the reported innovative activity. . . . [A]lmost all of the technical change studies examine the history of only one or two firms."

be expressed as output per unit of factor input ($y = Y/F$) times number of plants (N), times factor input per plant ($f = F/N$). In discrete growth terms:

$$\begin{array}{llllllll} (8) & \Delta Y/Y_{t-1} & = & (\Delta y/y_{t-1})(\omega_1) & + & (\Delta f/f_{t-1})(\omega_2) & + & (\Delta N/N_{t-1})(\omega_3) \\ & G & = & G_1 & + & G_2 & + & G_3. \end{array}$$

The first term on the right side of equation 8 reflects productivity growth, the second term reflects changes in the average scale of operations (measured by factor use), and the third term reflects net entry. The weights ω_1, ω_2, and ω_3 are averages of all possible variants of the identity and reflect the fact that the equation is in discrete terms; each weight will be close to one. The identity represented by equation 8 reveals not only whether output expansion has come mainly from productivity growth, but also whether productivity changes have been accompanied by changes in scale or net entry.

Annual data on all plants with at least 10 workers were available for three of the countries in the ICPT project: Chile 1979–85, Colombia 1977–87, and Morocco 1984–87. Using plant identification codes, it was possible to identify plants entering and exiting each of these data bases and to thereby implement equation 8. (Ideally entry and exit would reflect the births and deaths of plants, respectively, but in practice they also reflect crossings of the 10-worker threshold.)

Letting F be number of workers so that G_1 measures growth in labor productivity, implementation of equation 8 resulted in the figures reported in table 1. We did not observe capital stock figures for all plants in all countries, so it was not possible to construct broader indexes of factor use. Broader productivity measures have yielded similar results. For example, Liu (1991) finds evidence in the Chile data that labor productivity and multifactor productivity measures are closely related.

In table 1 values are given for all plants in the manufacturing sector and for three broad subgroups: exportable producers, importable producers, and nontradable producers. A firm is classified as producing an exportable product if its three-digit industry exports more than 25 percent of its output, on average. A firm is classified as producing an importable product if its three-digit industry exports less than 25 percent of its output, but at least 25 percent of the domestic market for this industry's good is supplied by imports. All other firms are classified as nontradable producers. Year-to-year fluctuations in our decomposition probably largely reflect capacity utilization effects. Hence, to get an overview of the long-run significance of each effect, we begin with an intertemporal average of each component value for each country.

There are two striking features of table 1. The first is that entry and exit (reflected by changes in the number of plants) are quite significant, which implies that there are high returns to improving our understanding of turnover processes. The second is that output expansion appears to occur through very different mechanisms in the different economies, even though the reported statistics are averages spanning at least four years. In Chile and Morocco adjust

Table 1. *Decomposition of Output Growth in the Manufacturing Sector in Chile, Colombia, and Morocco*
(average annual growth rate)

Country, period, and plant category	*Output (G)*	*Labor productivity (G_1)*	*Scale (G_2)*	*Number of plants (G_3)*
Chile (1979–85)				
All plants	0.014	0.044	0.018	−0.049
Exportables	0.085	0.093	0.034	−0.042
Importables	0.001	0.036	0.014	−0.049
Nontraded	−0.018	0.013	0.016	−0.047
Colombia (1977–87)				
All plants	0.043	0.046	0.006	−0.009
Exportables	0.033	0.034	0.009	−0.010
Importables	0.056	0.058	0.005	−0.008
Nontraded	0.038	0.041	0.007	−0.010
Morocco (1984–87)				
All plants	0.046	−0.038	0.009	0.075
Exportables	0.026	−0.105	0.042	0.089
Importables	0.083	−0.001	0.013	0.071
Nontraded	0.049	0.018	−0.038	0.069

Source: Author's calculations based on World Bank data.

ment comes largely from entry and exit, while in Colombia adjustments in the productivity of incumbent plants appear more important. Sector-specific differences in the nature of adjustment are also apparent: tradables appear to accomplish more adjustment through entry and exit than nontradables in Colombia and Morocco. These results suggest that behavioral models emphasizing sunk costs and uncertainty hold promise.

Decomposition of Growth Factors

Table 1 obscures the fact that entering, exiting, and incumbent plants probably differ systematically in size and productivity. Documenting these differences should help us to understand the influence of plant heterogeneity on measured TFP. To this end one may further decompose each element on the right side of equation 8. First, the productivity growth index, G_1, reflects three influences:

$$(9) \quad \begin{aligned} \Delta y/y &= \Delta y_c \bar{\alpha}_c/y + \Delta\alpha_c[\bar{y}_c - 1/2(y_d + y_b)]/y + (y_b - y_d)(1 - \bar{\alpha}_c)/y \\ G_1/\omega_1 &= G_{11}/\omega_1 + G_{12}/\omega_1 + G_{13}/\omega_1. \end{aligned}$$

Here α_c is the proportion of total factor use accounted for by plants that were in the industry both last period and this period (hereafter "continuing" or "incumbent" plants), and y_c is average productivity among these plants. (A bar above a variable indicates an average of last period's and this period's value.) Similarly, y_b is productivity among plants that have entered the industry this period, and y_d is productivity among plants that were in the industry last period, but exited this period. Hence the first term (G_{11}) indicates what portion of productivity

growth is due to productivity improvements among incumbents, the second term (G_{12}) indicates how changes in the market share of incumbent plants influence productivity (hereafter, the "net entry effect"), and the last term (G_{13}) reflects any improvement in productivity from replacing exiting plants with entering plants (hereafter the "turnover effect").

If entering plants are more productive than exiting plants, G_{13} will be positive, reflecting desirable turnover effects. However, increases in the net entry rate cause the share of employment among incumbent plants to fall ($\Delta\alpha_c < 0$) so, if productivity is higher among incumbents than the *average* productivity among entering and exiting plants, overall productivity growth may be dampened by increases in entry.

Growth in scale, G_2 in equation 8, can be similarly decomposed. Defining $\lambda_c = N^c/N$ as the number of incumbent plants divided by the total number of plants, then

$$(10)\quad \begin{aligned} \Delta f/f &= \Delta f_c\lambda_c/f + \Delta\lambda_c[f_c - 1/2(f_d + f_b)]/f + (f_b - f_d)(1 - \lambda_c)/f \\ G_2/\omega_2 &= G_{21}/\omega_2 + G_{22}/\omega_2 + G_{23}/\omega_2. \end{aligned}$$

Here G_{21} reflects expansion in plant size (that is, factor use) among incumbents weighted by incumbents' market share. Size differences among incumbents, entrants, and exiting plants are reflected by G_{22} and G_{23}. The former, the net entry effect, is the gap between incumbent size and the average size among entering and dying plants, all weighted by the change in incumbent share. The latter, the turnover effect, is the difference in average size between entering and dying plants, weighted by the share of nonincumbents.

The last term in equation 8 (G_3) represents the effect of net entry on expansion. It can be decomposed into the difference between entry and exit rates:

$$(11)\quad \begin{aligned} \Delta N/N &= N^b/N - N^d/N \\ G_3/\omega_3 &= G_{31}/\omega_3 - G_{32}/\omega_3. \end{aligned}$$

Aggregate effects of turnover on growth. Values of elements on the right side of equations 9, 10, and 11 are presented in table 2. Again, to approximate long run values, these are averages over annual growth rates. Plant turnover plays a significant role in determining growth rates. If there were no entry and exit growth would simply reflect productivity increases and expansion among incumbents ($G_{11} + G_{21}$). But this figure is typically very different from *realized* total growth figures (G, in table 1). For example, output grew by an annual average of 1.4 percent in the Chilean manufacturing sector; with zero net exit this figure would have been 2.4 percent (0.030 − 0.006 = 0.024). Similarly, net entry accounts for 2 percent of Morocco's 4.6 percent average annual output growth.

Decomposition of growth in labor productivity. The figures in tables 1 and 2 also suggest productivity differences among incumbent, entering, and exiting plants. In Chile and Morocco the net entry (G_{12}) and turnover (G_{13}) compo

Table 2. *Detailed Growth Decomposition for the Manufacturing Sector in Chile, Colombia, and Morocco*
(average annual growth rates)

Country, period, and plant category	Labor productivity			Scale			Number of plants	
	Incumbents (G_{11})	Net entry (G_{12})	Turnover (G_{13})	Incumbents (G_{21})	Net entry (G_{22})	Turnover (G_{23})	Gross entry (G_{31})	Gross exit (G_{32})
Chile (1979–85)								
All plants	0.030	0.012	0.002	−0.006	0.018	0.006	0.062	0.111
Exportables	0.078	−0.002	0.016	−0.003	0.008	0.029	0.125	0.167
Importables	0.030	0.001	0.006	−0.002	0.014	0.003	0.073	0.123
Nontraded	−0.001	0.011	0.002	−0.009	0.016	0.009	0.065	0.112
Colombia (1977–87)								
All plants	0.037	0.004	0.006	−0.004	0.009	0.002	0.162	0.172
Exportables	0.032	−0.002	0.005	−0.005	0.007	0.007	0.175	0.185
Importables	0.039	0.008	0.012	−0.006	0.007	0.004	0.184	0.192
Nontraded	0.037	0.003	0.001	−0.003	0.012	−0.002	0.136	0.146
Morocco (1984–87)								
All plants	−0.021	−0.017	0.001	0.041	−0.039	0.007	0.135	0.060
Exportables	−0.072	−0.031	−0.001	0.058	−0.039	0.022	0.181	0.093
Importables	−0.010	−0.006	0.015	0.051	−0.016	−0.022	0.171	0.100
Nontraded	0.028	−0.013	0.002	0.004	−0.028	−0.014	0.155	0.086

Source: Author's calculations based on World Bank data.

nents of labor productivity are both generally non-zero. In particular Chile's net exit increases the market share of incumbents, and this improves productivity. Indeed, among Chilean importables and nontradables, this is the *main* component of productivity change (compare G_{12} and G_1). Net entry does the opposite in Morocco.

Differences in productivity between entering and exiting plants (G_{13}) generally account for changes in sectoral aggregates that are significant, but these changes are smaller than those resulting from productivity gaps between incumbents and others. Exiting plants tend to be less productive than the entering plants that displace them. Part of the productivity difference between incumbent plants and others results from the fact that the former are relatively large and more capital intensive, but the result is robust to multifactor productivity measures (Liu 1991; Backinezos 1991).

Decomposition of growth in scale. It appears that scale heterogeneity is also significant. Incumbents are much larger than either entering or exiting plants (Roberts 1989; Tybout 1989), so the net exit that takes place in Chile significantly increases average plant size (G_{22}), even though incumbent plants are shrinking ($G_{21} < 0$). The same logic works in reverse for Morocco, when rapid net entry takes place. Thus, to the extent that scale economies matter, the procyclical tendencies of factor productivity are likely to be dampened by turnover. However, Liu (1991) finds that learning-by-doing is relatively rapid among entrants that survive.

Comparison with U.S. patterns. It is surprising that the gross entry and exit rates of all three countries match or exceed those found in the United States. Hence the popular view that institutional barriers to entry and exit are relatively important in developing countries is not borne out. This is not immediately apparent from table 2; it can be inferred by comparing the five-year entry/exit patterns reported in Dunne and others (1988) with cohort survival figures reported in Roberts (1989), Tybout (1989), and Haddad and others (1990). Moreover plants that enter and exit in these countries are as large relative to the industrywide average plant size as they are in the United States.

Trade, Entry, and Exit

All of the above points in the direction of productivity analysis based on changing populations of heterogeneous plants. An emerging analytical literature provides some theoretical underpinnings (Jovanovic 1982; Pakes and Ericson 1987; Jovanovic and Lach 1989; Lambson 1989), but most of the models are too abstract to lend themselves to empirical implementation (see, however, Olley and Pakes 1990, Ericson and Pakes 1987, and Dunne and others 1989). This literature does suggest that we might improve our understanding of growth processes by devoting more empirical attention to the entry, exit, and growth processes; cohort-specific cost or production functions; and learning curves.

Evidence on the first issue will help us understand the composition of an industry at each point in time, whereas evidence on the latter two issues will help us map alternative compositions into associated sectorwide productivity levels. The functions representing entry, exit, growth, and production or cost should be time-dependent and potentially sensitive to regime changes.

Several stylized facts concerning growth, entry, and exit are well established. Specifically, as new plants mature, their probability of failure drops, and their size increases toward industry norms. For recent studies of the United States see Dunne and others (1989) and Evans (1987); for ICPT project countries see Haddad and others (1990), Roberts (1989), and Tybout (1989). However, there is relatively little in the literature documenting the effects of trade policy on these processes. To provide some preliminary results, regressions linking entry and exit with industrial output growth and import penetration rates were attempted for several of the ICPT project countries (Tybout 1989; Roberts 1989). These regressions were done using annual data on three-digit industries, controlling for fixed industry effects and time effects with dummies. Hence the coefficients on import penetration rates were identified with temporal fluctuations, not cross-industry contrasts.

Several findings emerged. First, output growth is positively correlated with entry but does not correlate significantly with exit. The latter depends heavily on time dummies, thus suggesting that macro phenomena like high interest rates are more important than product market fluctuations in causing failures. Second, controlling for fixed industry effects, macro conditions, and output growth, fluctuations in import penetration do not correlate significantly with entry and exit patterns. This could mean that our regression model fails to capture the relevant dynamics, or it could mean that all significant effects of import penetration are picked up through industrywide fixed effects and output fluctuations. Third, industry dummies are generally significant, so to the extent that liberalization changes patterns of specialization, it *is* likely to change economywide rates of turnover. Finally, many coefficients are sensitive to the sample period chosen, thus possibly reflecting changing levels of uncertainty.

Pursuing the issue further, Backinezos (1991) used the Colombian data panel to fit industry-specific cost functions jointly with a probit function that controlled for selectivity due to exit. She found that in most sectors temporal variations in the import penetration rate were not associated with the probability of failure. Among those sectors where significant correlations were found, there were roughly as many positive correlations as negative ones. In short, although much remains to be done in exploring the dynamics of plant turnover, we have not yet uncovered systematic relations between entry and exit patterns and exposure to international competition.

The Size Distribution of Plants

Although contemporaneous correlations reveal no strong association between import penetration and entry and exit patterns, foreign competition might still

affect the size distribution of plants by inducing size adjustments among incum bents or by inspiring entry or exit with a lag. In turn, given the presence of scal economies, these shifts in size distributions may affect industrial productivity With these links in mind, we next turn to empirical evidence on the relatio between trade regimes and the size distribution of plants.

As Berry (1992) notes, earlier studies have found that larger plants are mor likely to be exporters. Similarly, it is sometimes conjectured that imports com pete more directly with large plants, given the nature of their product lines. Bu it is difficult to find studies that correlate trade policy with the size distributio of plants in developing countries. Econometric studies of this issue are also rar for developed countries; Baldwin and Gorecki (1983) is the only example I ar aware of. The ICPT project provides three exceptions.

Roberts and Tybout (1991) compare the plant size distributions in Chile an Colombia, industry by industry, and relate contrasts to associated differences i trade regime. To summarize the distribution of plant sizes for industry i, countr j, and year t, they rank plants by ascending employment level and find th employment levels of plants at the 10th, 25th, 50th, 75th, and 90th percentile: This generates five size measures: $\ln(\text{EMP}k_{ijt})$ = logarithm of the kth percentil of the employment size distribution ($k = 10, 25, 50, 75, 90$).

Each of these measures is regressed on proxies for various types of deman determinants. This can be done by exploiting either temporal variation, cros country variation, or both. Given that the former requires modelling of th dynamics of adjustment and that the latter accounts for most of the variation i the data, we summarize the cross-country results here. These are based o industry- and country-specific averages of annual values for each variable, an should approximate long-run relations. Letting bars above variables denot temporal averages, the estimated regressions are:

$$(12)\qquad \overline{\text{EMP}}k_{ij} = \beta_1\overline{\ln Y}_{ij} + \beta_2\overline{\text{ERP}}_{ij} + \beta_3\overline{\text{TUR}}_{ij} + \beta_4\overline{\text{TUR}}_{ij}\overline{\text{ERP}}_{ij} + \beta_5\overline{\text{TUR}}_{ij}\overline{\ln Y}_{ij} + \lambda_i + \mu_j + \epsilon_{ij}$$

where $\ln Y_{ijt}$ is the log of real industry output, ERP_{ijt} is the effective rate c protection, and TUR_{ijt} is the turnover rate. The turnover rate is the sum of th industry's entry and exit rates (that is, $G_{31}/\omega_3 + G_{32}/\omega_3$). These rates ar averaged across all years for each industry in each country to get a long-ru value that is specific to each industry in each country. Here the log of re: industry output proxies total demand for the industry's output, the effective rat of protection proxies protection from international markets, and the turnove rate proxies the ease of entry and exit. As suggested by many models of trad with imperfect competition, the sensitivity of size distributions to demand shif should depend upon the ease of entry and exit. The turnover variable is ther fore interacted with the log of real industry output and the effective rate (protection. Finally, to control for industry-specific technology effects an country-specific conditions, represented by λ and μ, respectively, industry an country dummies are included.

The results indicate that higher effective protection rates are associated with larger plant sizes, especially at the low end of the size distribution. This suggests that demand contraction, factor market effects, and other forces associated with increased import competition dominate any expansionary forces coming from higher demand elasticities in open economies; Baldwin and Gorecki (1983) found similar effects in Canadian data. Also the negative correlation of size and trade exposure is more substantial in low-turnover industries. This is consistent with the theoretical result that more size adjustment occurs when exit is not easy (Rodrik 1988; Buffie and Spiller 1986). Alternatively, the results may simply mean that the discipline of foreign competition matters more in industries where the discipline of potential entry is less important. All of these results hold up if size is measured with output instead of employment. They also hold up if the effective rate of protection is replaced with import penetration rates and export ratios—in fact this strengthens them. In regressions based on time series (rather than cross-country) variation in the data, however, no clear pattern emerges. This probably reflects the inability of the model to capture the dynamics of adjustment.

In a related study Tybout and others (1991) compare the 1967 and 1979 Chilean industrial censuses and ask whether sectors that underwent relatively large reductions in effective protection between the census years showed distinctive shifts in their size distribution. They look at the cross-industry Spearman rank correlation between changes in an effective protection measure and changes in the various size percentiles. This exercise is very similar to the Roberts and Tybout (1991) regressions, except here variation is examined across time rather than across countries.

When they measure size with employment, Tybout and others (1991) weakly confirm Roberts and Tybout's finding that higher levels of protection are associated with larger plant sizes, controlling for industry-specific effects and the state of the macroeconomy. However, when size is measured with output or value-added, Tybout and others find a *negative* weakly significant association between protection and size for the lower percentiles and a positive insignificant association for the higher percentiles. This implies that labor productivity among small plants tends to improve most in those industries undergoing the most dramatic reductions in effective protection.

Tracking Moroccan plants through time, Dutz (1991) also finds that import expansion caused by the relaxation of quotas was associated with contractions in plant size. But unlike the studies summarized above, he finds that small firms contracted proportionately more than large firms. So it is possible that efficiency improved with liberalization because of scale economy exploitation in Morocco. Much remains to be done in determining whether the findings of these three studies generalize to other countries and liberalization episodes, but they do cast doubt on the popular conjecture that opening an economy leads to efficiency gains through the exploitation of plant-level scale economies.

Scale Economies

To link changes in the size distribution of plants with productive efficiency, it is necessary to have some sense of the importance of plant-level scale economies. After reviewing various problems associated with econometric estimates of the returns to scale—due especially to measurement error and unobservable management effects—Tybout and Westbrook (1991) settle on a "method of moments" estimator that predicts changes in output with changes in input vectors over long periods, using instruments to reduce the measurement error in capital stock growth rates. They find econometric evidence that their estimator works better than more naive variants but also that very few Chilean industries show returns to scale significantly different from unity. Similar findings emerged from other project output that focused less on methodological issues (Tybout and others 1991). Hence, although external scale economies resulting from infrastructure and learning spillovers may be important, project participants could find little empirical support for promoting "bigness" at the plant level.

Plant-Specific Technologies

In the preceding analysis we have been looking for links between trade policy and productivity that are based on plant heterogeneity. Thus far we have been unable to establish a strong link between trade patterns and turnover. We have also seen that, if anything, increases in trade exposure appear to *contract* plants in terms of employment. So if we are to muster evidence that trade liberalization improves productive efficiency through heterogeneity effects, it must come from a change in the technical efficiency of incumbent plants or from changes in the types of plants that enter and exit. This section argues that there is some evidence of the former, but as yet the latter has not been researched.

Perhaps the most popular approach to documenting plant-specific technologies is based on efficiency frontiers. The notion is that if the frontier production technology $Y = f(\mathbf{v},t)$ represents the maximum output achievable with the input vector $\mathbf{v}$ at time t, then production observed at the ith plant will fall short of the frontier by some amount $e_i = f(\mathbf{v}_i,t) - Y_i$. If the technology $f(\cdot)$ can be estimated, then one can obtain not only a measure of returns to scale, but also a set of plant-level inefficiency indexes, e_i. A variety of approaches to estimating $f(\cdot)$ exist; partial surveys may be found in Schmidt (1985), van den Broek and others (1980), and Forsund and others (1980).

Pack (1988) notes that among studies of countries pursuing import substitution, large intra-industry differences in productivity are common, as are low average productivity levels relative to the frontier technology (Handoussa and others 1986; Page 1984; Pack 1987). Pack (p. 363) speculates that similar work "in export-oriented countries would reveal considerably smaller intra-industry variation in TFP as well as a better average *level* of TFP" but stresses that all evidence points in the direction of one-time improvements in the *level* of TFP

with trade liberalization, rather than improvements in the long-run rate of TFP growth. Studies by Chen and Tang (1987) and Handoussa and others (1986) are consistent with this conjecture. Havyrylshyn (1990) is less cautious in his assessment. After reviewing several studies applying efficiency frontiers to developing countries (Pitt and Lee 1981; Nishimizu and Page 1991; Page 1984), he concludes that the studies generally "found strong empirical evidence of a positive effect of trade policy liberalization" (p. 16).

Further support for the conjecture that trade liberalization improves technical efficiency may be found in Tybout and others (1991), which compares Chilean industrial census data from an import-substituting period (1967) and an outward-oriented period (1979). Tests for the effects of liberalization are based on the following observation: if exposure to foreign competition improves average efficiency and causes plants below minimum efficient scale to exit, then the estimated production technology should show an increase in its intercept, a reduction in estimated returns to scale, and a reduction in residual variation. Rather than use frontier production functions to look for these patterns, econometric attention is devoted to correcting for measurement error in capital stocks and missing data (Tybout 1992). It is found that, although no more than half of the three-digit industries registered overall productivity gains, the ones that showed the most improvement tended to be those that underwent the largest reductions in effective protection.

IV. Concluding Remarks

In the past decade there has been a growing realization that traditional Tornqvist indexes of productivity growth actually pick up much more than innovation, scale economies, and movement to the efficient frontier. One conclusion that emerges with force from recent work is that we should approach the reported figures with skepticism. Problems of measurement error, disequilibria, and aggregation bias can easily create the illusion of trends and correlations that have no basis in the economic processes we hope to capture.

These observations are dispiriting, but they have helped to inspire new ways of thinking about productivity growth and new approaches to looking for it. This article has reviewed two new directions. The first is an attempt to salvage traditional residual-based calculations by correcting for scale economies, adjustment costs, or noncompetitive pricing. Although the methodology still suffers from significant measurement problems and aggregation bias, it gives some sense for the robustness of productivity growth series to violations of the litany of assumptions. It also has the obvious advantage of being based on easily accessible data.

The second new direction discussed herein concerns the role of plant heterogeneity in shaping sectoral productivity growth. Except for work on frontier production functions, this strand of the literature is in its infancy, so many of the techniques are still quite rudimentary. Nonetheless they give a crude sense for

the importance of entry, exit, and heterogeneity in shaping productivity growth patterns, as well as some specifics on the nature of aggregation bias in industry-level studies.

Throughout the discussion of new approaches to productivity measurement, performance measures have been examined for correlation with crude indexes of trade regime. In view of the diverse, ambiguous theoretical literature on the link between trade and productivity, it is not surprising that stable, predictable correlations have not emerged. Nonetheless in some countries and during some subperiods there is some association between trade flow patterns and indexes of productivity growth at the industry level, even after correcting for several measurement problems. Also we may tentatively conclude that the effects of trade regimes on productivity growth are interrelated with market concentration, although the nature of this association is itself unstable.

The lack of stable correlations in sectoral and industry-level data is matched by a surprising diversity in the processes of entry, exit, and scale adjustment. In some economies a good deal of output fluctuation appears to come from the creation and death of plants, whereas in others the incumbent plants are what matter. (Cross-country differences in institutions and the degree of uncertainty are possible explanations.) Given that there is considerable variation in size and labor productivity across these plant types, this is one reason the sectoral and industry-level analyses differ. The ICPT project has focused on linking entry, exit, and adjustments in scale and technical efficiency with exposure to trade regime. Thus far it appears that exposure to increased foreign competition is not closely linked with entry patterns, tends to induce reductions in plant size, and may cause some improvements in technical efficiency.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Abbott, Thomas, Zvi Griliches, and Jerry Hausman. 1989. "Short-run Movements in Productivity: Market Power versus Capacity Utilization." Harvard University, Department of Economics, Cambridge, Mass. Processed.

Backinezos, Constantina. 1991. "The Effects of Trade Liberalization on Sectoral Productivity in Colombian Manufacturing." Ph.D. dissertation, Pennsylvania State University, Department of Economics, University Park. Processed.

Balassa, Bela. 1985. "Exports, Policy Choices, and Economic Growth in Developing Countries after the 1973 Oil Shock." *Journal of Development Economics* 18 (May/June): 23–35.

Baldwin, J. R., and P. K. Gorecki. 1983. "Trade, Tariffs, and Relative Plant Scale in Canadian Manufacturing Industries: 1970–79." Discussion paper 232, Economic Council of Canada, Kingston, Ontario. Processed.

Baldwin, Richard. 1989. "Sunk-Cost Hysteresis." NBER Working Paper 2911, National Bureau of Economic Research, Cambridge, Mass. Processed.

Baldwin, Richard, and Paul Krugman. 1989. "Persistent Trade Effects of Large Exchange Rate Shocks." *Quarterly Journal of Economics* 104 (November): 635–54.

Berry, Albert. 1992. "Firm or Plant Size in the Analysis of Trade and Development." In Gerald Helleiner, ed., *Trade Policy, Industrialization and Development: New Perspectives*. Oxford: Clarendon Press.

Buffie, Edward, and Pablo Spiller. 1986. "Trade Liberalization in Oligopolistic Industries." *Journal of International Economics* 20 (February): 65–81.

Caballero, Ricardo, and Richard Lyons. 1989a. "Internal versus External Economies in European Industry." Discussion Paper 427, Columbia University, Department of Economics, New York. Processed.

——. 1989b. "The Role of External Economies in U.S. Manufacturing." Discussion Paper 431, Columbia University, Department of Economics, New York. Processed.

——. 1990. "Externalities and Cyclical Factor Productivity." Discussion Paper 467, Columbia University, Department of Economics, New York. Processed.

ıen, Tain-Jy, and De-piao Tang. 1987. "Comparing Technical Efficiency between Import-Substituting and Export-Oriented Foreign Firms in a Developing Country." *Journal of Development Economics* 26: 277–89.

enery, Hollis, Sherman Robinson, and Moshe Syrquin. 1986. *Industrialization and Growth: A Comparative Study*. New York: Oxford University Press.

rden, Max. 1974. *Trade Policy and Economic Welfare*. Oxford: Clarendon Press.

rnwell, Christopher, Peter Schmidt, and Robin Sickles. 1990. "Production Frontiers with Cross-Sectional and Time Series Variation in Efficiency Levels." *Journal of Econometrics* 46: 185–200.

xit, Avinash. 1989a. "Hysteresis, Import Penetration, and Exchange Rate Pass-Through." *Quarterly Journal of Economics* 104 (2, May): 207–28.

——. 1989b. "Entry and Exit Decisions under Uncertainty." *Journal of Political Economy* 97 (June): 620–38.

mowitz, Ian, R. Glenn Hubbard, and Bruce Peterson. 1986. "Business Cycles and the Relationship between Concentration and Price-Cost Margins." *Rand Journal of Economics* 17 (Spring): 1–17.

——. 1988. "Market Structure and Cyclical Fluctuations in U.S. Manufacturing." *Review of Economics and Statistics* 70 (February): 55–66.

nne, Timothy, Mark Roberts, and Larry Samuelson. 1988. "Patterns of Firm Entry and Exit in U.S. Manufacturing Industries." *Rand Journal of Economics* 19 (Winter): 495–515.

——. 1989. "The Growth and Failure of U.S. Manufacturing Plants." *Quarterly Journal of Economics* 104 (November):671–98.

tz, Mark A. 1991. "Firm Output Adjustment to Trade Liberalization: Theory with Application to the Moroccan Experience." PRE Working Paper 602, World Bank, Country Economics Department, Washington, D.C. Processed.

wards, Sebastian. 1989. "Openness, Outward Orientation, Trade Liberalization, and Economic Performance in Developing Countries." PRE Working Paper 191, World Bank, Country Economics Department, Washington, D.C. Processed.

cson, Richard, and Ariel Pakes. 1987. "Empirical Implications of Alternative Models of Firm Dynamics." SSRI Workshop Series 8803, University of Wisconsin—Madison, Department of Economics. Processed.

Evans, Paul D. 1987. "Tests of Alternative Theories of Firm Growth." *Journal of Political Economy* 95: 657–74.

Forsund, Finn, C. A. Knox Lovell, and Peter Schmidt. 1980. "A Survey of Frontier Production Functions and of Their Relationship to Efficiency Measurement." *Journal of Econometrics* 13: 5–25.

Gomulka, Stanislaw. 1976. "Do New Factories Embody Best Practice Technology?—New Evidence." *Economic Journal* 86 (December): 859–63.

Grossman, Gene, and Elhanan Helpman. 1989. "Endogenous Product Cycles." working Paper 10-89, Foerder Institute for Economic Research, Tel Aviv, Israel.

———. 1990. "Comparative Advantage and Long-Run Growth." *American Economic Review* 80 (September): 796–815.

Haddad, Mona, Brendan Horton, and Jaime de Melo. 1990. "Entry, Exit, Competition, and Productivity in the Moroccan Industrial Sector." World Bank, Country Economics Department, Washington, D.C. Processed.

Hall, Robert. 1986. "Market Structure and Macroeconomic Fluctuations." *Brookings Papers on Economic Activity* 2: 285–322.

———. 1988a. "The Relation between Price and Marginal Cost in U.S. Industry." *Journal of Political Economy* 96 (October): 921–47.

———. 1988b. "Increasing Returns: Theory and Measurement With Industry Data." NBER Program on Economic Fluctuations, National Bureau of Economic Research, Cambridge, Mass. Processed.

Handoussa, Heba, Mieko Nishimizu, and John Page. 1986. "Productivity Change in Egyptian Public Sector Industries after the 'Opening.'" *Journal of Development Economics* 20: 53–74.

Harrison, Ann E. 1990. "Productivity, Imperfect Competition, and Trade Liberalization in Côte d'Ivoire." PRE Working Paper 451, World Bank, Country Economics Department, Washington, D.C. Processed.

Hart, Oliver. 1983. "The Market Mechanism as an Incentive Structure." *Bell Journal of Economics* 14 (Autumn): 366–82.

Havyrlyshyn, Oli. 1990. "Trade Policy and Productivity Gains in Developing Countries: A Survey of the Literature." *The World Bank Research Observer* 5 (1): 1–24.

Helpman, Elhanan. 1990. "The Noncompetitive Theory of International Trade and Trade Policy." In World Bank, *Proceedings of the World Bank Annual Conference on Development Economics, 1989*. Washington, D.C.

Jorgenson, Dale, Frank Gallop, and Barbara Fraumeni. 1987. *Productivity and U.S. Economic Growth*. Cambridge, Mass.: Harvard University Press.

Jovanovic, Boyan. 1982. "Selection and the Evolution of Industry." *Econometrica*. 50 (May): 649–70.

Jovanovic, Boyan, and Saul Lach. 1989. "Entry, Exit, and Diffusion with Learning by Doing." *American Economic Review* 79 (4, September): 690–99.

Krueger, Anne O., and Baran Tuncer. 1982. "Growth of Factor Productivity in Turkish Manufacturing Industries." *Journal of Development Economics* 11: 307–25.

Krugman, Paul. 1985. "The Narrow Moving Band, the Dutch Disease, and the Competitive Consequences of Mrs. Thatcher: Notes on Trade in the Presence of Dynamic Scale Economies." MIT, Department of Economics, Cambridge, Mass. Processed.

———. 1986. "New Trade Theories and the Less-Developed Countries." In G. Calvo,

ed., *Debt, Stabilization and Development: Essays in Honor of Carlos Diaz-Alejandro*. New York: Basil Blackwell.

Lambson, Val Eugene. 1989. "Industry Evolution with Sunk Costs and Uncertain Market Conditions." SSRI Working Paper 8904, University of Wisconsin—Madison, Department of Economics. Processed.

Levinsohn, James. 1991. "Testing the Imports-as-Market-Discipline Hypothesis." NBER Working Paper 3657, National Bureau of Economic Research, Cambridge, Mass. Processed.

Liu, Lili. 1991. "Entry-Exit, Learning, and Productivity Change: Evidence from Chile." PRE Working Paper 769, World Bank, Country Economics Department, Washington, D.C. Processed.

Lucas, Robert E., Jr. 1988. "On the Mechanics of Economic Development." *Journal of Monetary Economics* 22 (July): 3–42.

Martin, John P., and John M. Page, Jr. 1983. "Impact of Subsidies on LDC Industry: Theory and an Empirical Test." *Review of Economics and Statistics* 65: 608–17.

Nelson, Richard R. 1981. "Research on Productivity Growth and Productivity Growth and Productivity Differences: Dead Ends and New Directions." Institution for Social and Policy Studies Working Paper 844, Yale University, New Haven, Conn. Processed.

Nishimizu, Mieko, and Sherman Robinson. 1984. "Trade Policies and Productivity Change in Semi-Industrialized Countries." *Journal of Development Economics* 16 (September/October): 177–206.

Nishimizu, Mieko, and John Page. 1991. "Trade Policy, Market Orientation, and Productivity Change in Industry." In Jaime de Melo and Andre Sapir, eds., *Trade Theory and Economic Reform: North, South and East—Essays in Honor of Bela Balassa*. Cambridge: Basil Blackwell.

Olley, G. Steven, and Ariel Pakes. 1990. "Dynamic Behavioral Responses in Longitudinal Data Sets: Productivity in Telecommunications Equipment Industry." University of Wisconsin—Madison, Department of Economics. Processed.

Pack, Howard. 1987. *Productivity, Technology, and Industrial Development*. New York: Oxford University Press.

———. 1988. "Industrialization and Trade." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*. Vol. 1. Amsterdam: North-Holland.

———. 1992. "Learning and Productivity Change in Developing Countries." In G. Helleiner, ed., *Trade Policy, Industrialization and Development: New Perspectives*. Oxford: Clarendon.

Page, John. 1984. "Firm Size and Technical Efficiency: Application of Production Frontiers to Indian Survey Data." *Oxford Economic Papers* 32: 19–39.

Pakes, Ariel, and Richard Ericson. 1987. "An Alternative Theory of Firm Dynamics." Columbia University, Department of Economics, New York. Processed.

Pitt, M., and L. Lee. 1981. "The Measurement and Sources of Technical Inefficiency in the Indonesian Weaving Industry." *Journal of Development Economics* 9: 43–64.

Roberts, Mark. 1989. "The Structure of Production in Colombian Manufacturing Industries." World Bank, Country Economics Department, Washington, D.C. Processed.

Roberts, Mark, and James R. Tybout. 1991. "Size Rationalization and Trade Exposure in Developing Countries." In R. Baldwin, ed., *Empirical Studies of Commercial Policy*. University of Chicago Press for the NBER.

Rodrik, Dani. 1988. "Imperfect Competition, Scale Economies, and Trade Policy in Developing Countries." In R. Baldwin ed., *Trade Policy Issues and Empirical Analysis*. University of Chicago Press for the NBER.

———. 1992. "Closing the Technology Gap: Does Trade Liberalization Really Help?" In G. Helleiner, ed., *Trade Policy, Industrialization and Development: New Perspectives*. Oxford: Clarendon.

Scharfstein, David. 1988. "Product-market Competition and Managerial Slack." *Rand Journal* 19 (Spring): 147–55.

Schmidt, Peter. 1985. "Frontier Production Functions." *Econometric Reviews* 4: 289–328.

Shapiro, Matthew. 1987. "Measuring Market Power in U.S. Industry." NBER Working Paper 2212. National Bureau of Economic Research, Cambridge, Mass. Processed.

Stewart, Frances, and Ejaz Ghani. 1992. "Externalities, Development, and Trade." In G. Helleiner, ed., *Trade Policy, Industrialization and Development: New Perspectives*. Oxford: Clarendon.

Tybout, James R. 1989. "Entry, Exit, Competition, and Productivity in the Chilean Industrial Sector." World Bank, Country Economics Department, Washington, D.C. Processed.

———. 1991. "Researching the Trade-Productivity Link: New Directions." PRE Working Paper 638. World Bank, Country Economics Department, Washington, D.C. Processed.

———. 1992. "Making Noisy Data Sing: Estimating Production Technologies in Developing Countries." *Journal of Econometrics* (February/March).

Tybout, James R., Jaime de Melo, and Vittorio Corbo. 1991. "The Effects of Trade Reforms on Scale and Technical Efficiency: New Evidence from Chile." *Journal of International Economics* 31: 231–50.

Tybout, James, and M. Daniel Westbrook. 1991. "Estimating Returns to Scale with Large Imperfect Panels." PRE Working Paper 754. World Bank, Country Economics Department, Washington, D.C. Processed.

van den Broek, Julien, Finn Forsund, Lennary Hjalmarsson, and Winn. Meeusen. 1980. "On the Estimation of Deterministic and Stochastic Frontier Production Functions." *Journal of Econometrics* 13: 117–38.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 2: 213-231

Transportation Policy and Panterritorial Pricing in Africa

Mark Gersovitz

Many African countries interfere with the spatial pattern of agricultural prices and often mandate a spatially constant (panterritorial) price. Such policies are conceptually identical to the pricing of transport services. The pricing and transportation policies of the Ivoirien cotton parastatal provide a case study. The loss in producer surplus from raising current revenues by panterritorial pricing is compared with an optimal policy that minimizes producer loss. For current revenues, a switch to the optimal policy would provide only small gains. At higher revenue levels, however, the gains would increase. These policies are also contrasted with another suboptimal policy—full-cost pricing of transport with an export tax. By dispersing production, panterritorial pricing inflates the gains from transport projects. Transport investment and pricing reform are therefore assessed simultaneously. The number and location of the purchasing depots are also discussed.

Governments influence the price and quality of transport services, both through their investments in transport infrastructure and through a range of regulations. In rural areas these decisions affect the spatial pattern of agricultural production, with consequences for the efficiency of agriculture, the well-being of agriculturalists, and the revenues obtained from the agricultural sector by the government.

In many African countries decisions on the pricing of transport services are made implicitly as a consequence of government interventions in the marketing of agricultural produce. Often these governments establish public or semipublic agencies (parastatals) with monopsony control over the marketing of export crops. The ways in which these parastatals pay farmers at different locations for their crops are equivalent to a set of regulations on the pricing and quality of transport services. One commonly adopted component of parastatal policy is panterritorial pricing, in which all farmers are paid the same price regardless of

Mark Gersovitz is in the Department of Economics and the Center for Research on Economic Development at the University of Michigan. This research was supported by the World Bank under RPO 674-75 and by a grant from the Lynde and Harry Bradley Foundation to the Research Program in Development Studies at Princeton University. The author wishes to thank Esra Bennathan, Alice Galenson, Jeffrey Guttman, and Christina H. Paxson for their comments on earlier drafts.

where their produce is purchased. This is the practice of the Compagnie Ivoirienne pour le Developpement des Textiles (CIDT), the parastatal that is involved with cotton production and marketing in Côte d'Ivoire. The case of the CIDT illustrates some of the policies on rural transport in Africa.

This article sets out some practical diagnostic tests of the efficiency of (implicit) transport policies embodied in state marketing and discusses the implications of the policies for infrastructural investment. These tests are feasible using readily available data for many African countries and crops. Section I presents a conceptual framework for looking at rural transport and state marketing, by building on the work of Walters (1968) on the Ellet model. Gersovitz (1989) applies the Ellet model to some topics in optimal taxation and agricultural marketing. Then these models are used to organize some basic facts about the operations of the CIDT. Section II models the implications of panterritorial pricing for transport policy and applies the model to the activities of the CIDT. Section III gives special attention to choices about the location of purchasing depots used by parastatals and the consequences for the implicit price of transport. In principle there are a wide variety of options here, and the CIDT exemplifies only one possibility. The final section makes some concluding remarks.

I. The Spatial Pattern of Production and Marketing in the Export Sector

To understand the demand for rural transport generated by the production of a crop, one must understand where the crop is produced relative to where it is consumed or exported. The essence of agricultural production is, of course, the importance of land as an input and the consequent spatial dispersion of production. The volume and location of production are determined both by the area where the crop is cultivated (the extensive margin of cultivation) and by the intensity of production on land that is used to produce the crop (the intensive margin of cultivation). The boundaries of the region in which a particular crop is grown in a particular country may be determined in various ways: by the limits of the country's borders, by the limits of the land that is at all physically suitable for cultivation of the crop, or by the limits of the land on which it is economically profitable to grow the crop. The first two factors are exogenous to the transport policies of the government, whereas the last is highly influenced by them.

In practice whether one emphasizes changes in the intensive or extensive margins of cultivation in agricultural response to transport policy depends partially on the degree of aggregation in the information that is available. For individual farms, it may be possible to see changes in the extensive margin, whereas at a more aggregate level the boundaries of the region where production occurs may not change at all. In Africa information on the effect of transport policy on the agricultural response of individual farms is largely unavailable, either because governments do not collect it or because they treat it as confidential.

For practical reasons, therefore, the study of rural transport is largely restricted to data reported at a geographical level just below that of the region in which production occurs or to special studies of particular road projects. For the CIDT the data come from the geographical level termed *zones productrices*. Information on the CIDT is from Beenhakker and Bruzelius (1985); DCGT (1986, 1988); and CIDT (various years a, b, c).

The cotton sector of Côte d'Ivoire occupies about 188,000 square kilometers, split into about 56 zones that the CIDT uses for organizational and reporting purposes (map 1). It is bordered on the east, west, and north by the international boundaries of the country. To the south it is limited by the replacement of the savanna, which is agroclimatically suited to cotton, by the forest, which is not. It does not, therefore, seem that an important part of the response of production to decreases in transport costs would come from an expansion in the borders of the cotton region, as would be reflected in an addition of zones. Consequently, I assume that the number of zones is fixed when calculating the effects of changes in the (implicit) price of transport on production, on the benefits received by agricultural producers, and on government revenues.

It is possible, however, that the region of production could contract from its current boundaries, with zones dropping out of production entirely, if the effect of a change in policy on farmers is sufficiently adverse. The constant-elasticity-of-supply function of the simulations implies that a zone does not drop out as long as its zonal price is positive. If the number of zones were to change with changes in the parameters of the model, then it would be necessary to keep track of which zones are in production at any given time and to alter the summation signs in the following equations correspondingly.

Calculation of the effects of transport policy requires a calculation of the benefits received by agricultural producers and the revenues received by the government. The first step is a calculation of the level of production. For the i^{th} zone, production (Q_i) is given by

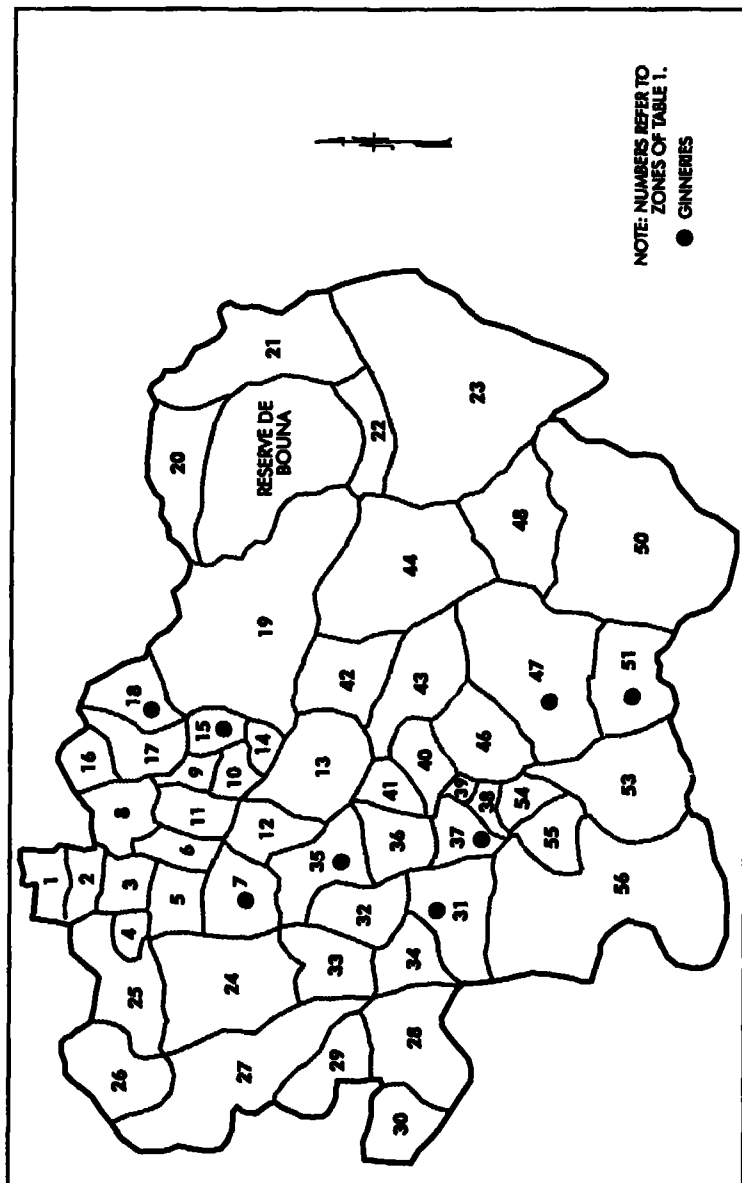
$$Q_i = \gamma_i S(p_i)/S(p_i^0) \tag{1}$$

where $S(\cdot)$ is a function that depends positively on the price (p_i) received by farmers in the zone for their production and p_i^0 is the initial (base case) price received by farmers for their output. When $p_i = p_i^0$, $Q_i = \gamma_i$, so γ_i is just the production of the zone at the initial price. The production response embodied in $S(\ )$ reflects the intensification of cultivation from adding inputs such as labor and fertilizer, bringing uncultivated land into cotton cultivation, and switching resources from the production of other crops.

Government revenues from the production in a particular zone (R_i) are therefore

$$R_i = [p^* - c^* - p_i - a^* t_i]\, Q_i \tag{2}$$

where p^* is the export price of the output in world markets; c^* is the costs per ton paid by the parastatal that do not vary with the region of production; a^* is the cost of transport per ton per kilometer paid by the parastatal; and t_i is the

Map 1. *The CIDT's Area of Operation in Côte d'Ivoire*

average distance from the zone to the port of export. The benefits to producers in a zone are measured by their producer surplus (Π_i):

$$(3) \qquad \Pi_i = p_i Q_i - \gamma_i \int_0^{S(p_i)} S^{-1}(x)dx / S(p_i^0).$$

The corresponding total values for the sector as a whole are simply

$$(4) \qquad R = \sum_{i=1}^{N} R_i$$

and

$$(5) \qquad \Pi = \sum_{i=1}^{N} \Pi_i$$

where N is the number of zones (56 for the CIDT).

The zones are listed in table 1, with values for the variables that vary by zone: zonal deliveries to ginneries, distance from zone to ginnery, distance from ginnery to port, and transport costs. The values of the price variables used to calculate the cost of transport and the producer surplus ratio are from 1982/83 and the quantity values are from 1987/88. To the extent that the relative prices differed between 1982/83 and 1987/88, this procedure introduces inaccuracies into the benchmarking of the model.

All calculations are in terms of ginned cotton fiber. One kilogram of seed cotton is about 0.419 kilogram of ginned cotton fiber. Therefore, because farmers produce seed cotton and sell it to the CIDT for processing at the ginneries, their output is measured in the modeling exercise as $\gamma_i = 0.419$ times their actual deliveries of seed cotton (column 1 of table 1). To adjust for differences in the weight of unginned and ginned cotton and for differences in the cost of transport at different stages, the following steps are used to calculate t_i: divide the distance from the zone to the ginnery (column 2) by 0.419, multiply the distance from the ginnery to the port (column 3) by 0.026 and divide by 0.067 to account for the differences in costs per kilometer per ginned cotton equivalent from the depot to the ginnery and from the ginnery to the port (table 2), and add the results from the preceding two steps to get an (adjusted) t_i. The value of a^* is then equal to CFAF0.067 per kilogram-kilometer.

One very weak test for efficiency in the transport policy implicit in these prices is that the return to selling cotton on the world market net of unavoidable costs paid by the CIDT, such as ginning (but not including transport costs), is greater than the costs to move cotton from the zone to the port ($p^* - c^* > a^* t_i$). Because it neglects the farmers' costs, the criterion could be met and production could still be socially undesirable. As the data in tables 1 and 2 show, this condition is easily met for all zones. The net return to selling cotton on the world market is CFAF350, while the largest value of the cost of transport from the zone to the port is CFAF76.3. Furthermore, with the government paying CFAF191 to

[able 1. *Deliveries, Distances, and Costs of Transporting Cotton n the Côte d'Ivoire, by Zone*

Zone		Deliveries[b] (kilograms of seed cotton)	Kilometers		Total cost of transport[e] (CFA francs per kilogram of ginned cotton)	Ratio of producer surpluses[f]
Number[a]	Name	(1)	From zone to ginnery[c] (2)	From ginnery to port[d] (3)	(4)	(5)
1	Tingrela	5,281,864	130	684	38.6	0.96
2	Bolona	5,857,804	117	684	36.5	0.97
3	Zanguinasso	3,929,054	85	684	31.4	1.00
4	Sanhala	3,747,105	99	684	33.6	0.99
5	Gbon	5,938,109	83	655	30.2	1.01
6	Kassere	6,111,095	85	669	31.0	1.00
7	Boundiali	6,182,039	86	586	29.1	1.01
8	Mbingue	13,936,334	165	530	40.2	0.96
9	Koni	8,690,719	57	547	23.4	1.04
10	Lataha	7,627,519	108	555	31.7	1.00
11	Niofouin	7,075,521	128	568	35.2	0.98
12	Sirasso	4,532,020	71	568	26.1	1.03
13	Dikodougou	6,062,180	123	542	33.8	0.99
14	Napie	4,739,669	93	516	28.4	1.02
15	Sinematiali	3,127,790	103	507	29.6	1.01
16	Nielle	11,360,400	71	621	27.5	1.02
17	Diawalla	8,099,950	57	621	25.3	1.03
18	Oungolo	9,070,930	55	597	24.2	1.04
19	Ferke	7,677,236	132	565	35.7	0.98
20	Tehini	619,376	280	564	59.4	0.86
21	Bouna	251,275	349	564	70.5	0.80
22	Nassian	588,744	295	361	56.6	0.87
23	Bondoukou	699,275	321	361	60.7	0.85
24	Madinani	2,677,229	292	504	59.8	0.86
25	Goulia	4,565,559	385	568	76.3	0.78
26	Tienko	4,420,004	384	504	74.5	0.79
27	Odienne	2,904,497	292	504	59.8	0.86
28	Touba	1,607,985	150	504	37.1	0.97
29	Borotou	1,591,761	159	504	38.5	0.96
30	Ouaninou	1,380,240	133	504	34.4	0.99
31	Seguela	3,651,850	20	504	16.3	1.08
32	Kani	4,935,878	104	459	28.6	1.02
33	Morondo	5,715,800	125	489	32.8	0.99
34	Worofla	2,377,068	46	504	20.5	1.06
35	Dianra	13,538,640	31	545	19.2	1.07
36	Sarhala	5,310,990	75	459	23.9	1.04
37	Mankono	6,369,784	21	459	15.3	1.09
38	Kounahiri	1,466,950	184	266	36.3	0.98
39	Foutounou	1,389,927	186	266	36.7	0.97
40	Tienigboue	4,887,478	61	459	21.7	1.05

Table 1. *continued*

Zone Number[a]	Zone Name	Deliveries[b] (kilograms of seed cotton) (1)	Kilometers: From zone to ginnery[c] (2)	Kilometers: From ginnery to port[d] (3)	Total cost of transport[e] (CFA francs per kilogram of ginned cotton) (4)	Ratio of producer surpluses (5)
41	Marandala	6,649,932	161	361	35.1	0.98
42	Niakara	4,106,491	147	361	32.9	0.99
43	Katiola	5,770,200	74	361	21.2	1.06
44	Dabakala	1,910,127	131	361	30.3	1.01
46	Beoumi	2,579,705	114	297	26.0	1.03
47	Bouake	904,872	32	361	14.5	1.09
48	M'Bahiakro	1,229,859	205	266	39.7	0.96
50	Bongouanou	2,602,899	257	266	48.0	0.92
51	Yamoussokro Nord	2,013,622	30	266	11.7	1.11
52	Yamoussokro Sud	1,555,122	—	—	—	—
53	Bouafle	5,463,219	125	266	26.9	1.03
54	Gohitafla	3,161,624	148	266	30.6	1.01
55	Zuenoula	5,634,192	134	266	28.3	1.02
56	Vavoua	7,899,487	129	404	31.2	1.00
	Mean	4,657,019	140	478	34.8	0.98
	Standard deviation	3,100,213	93	130	14.8	0.08
	Minimum	251,275	20	266	11.7	0.78
	Maximum	13,936,334	385	684	76.3	1.11

— Not available.

a. There are 53 zones reporting data. Data for zones 45, 49, and 52 are not reported for 1987/88.

b. Deliveries to ginneries are in (unginned) seed cotton 1987/88.

c. Weighted average of the distances to each ginnery to which deliveries were made in 1988 from a given zone, weighted by the fraction of total deliveries from the zone going to that ginnery.

d. Distance the cotton from a particular zone travels *after* it has been ginned. In a first step, distance to the port from a particular ginnery is a weighted average of the distances from the ginnery to the ports of Abidjan and San Pedro, weighted by the share of each port in the shipments of the particular ginnery. In the second step, the distance of the zone to the port is then the weighted average of the distances from the ginneries to the ports defined in step 1 to which the zone ships, weighted as in the calculations for distance from zone to ginnery. The ginnery at Dianra was assigned the same port shares as the ginnery at Mankono, and the ginnery at Seguela was assumed to ship all its cotton to San Pedro.

e. From farmgate to port via the ginnery. Calculation is based on one kilogram of seed cotton being equivalent to 0.419 kilogram of ginned cotton. See also the discussion in the text.

f. The ratio of producer surplus under optimal pricing to that under panterritorial pricing for an elasticity of supply of 0.67 and an R as in table 3.

Source: CIDT (various years c) for deliveries; unpublished data provided by the Caisse Centrale in Abidjan for distance from zone to ginnery; and Michelin (1989) and DCGT (1986, p. 24) for data to calculate weighted distance from ginnery to port. For ratio of producers' surplus calculations from model, as discussed in text.

farmers regardless of the zone, the net return to selling cotton on the wor[l]d market, taking into account the price received by farmers, is CFAF*159; thi[s]* exceeds the largest value of the cost of transport by CFAF*82.7. Thus the* government gains revenue from every single zone, although the revenue per kilogram is less from the more remote zones.

II. PANTERRITORIAL PRICING AND ITS ALTERNATIVES

Relatively remote places of production have relatively higher costs of transport to the place of ultimate consumption or export. Governments that buy farm output at the same price everywhere charge farmers in remote areas nothing for these higher costs. They may be said to subsidize fully the differential transport cost between remote and near-in farmers. Production is more dispersed than is efficient. Farmers may incur some transport costs in getting their output to the place of purchase, a cost of more or less significance, depending on the location of the government's buying depots, an issue that is given more attention in the next section. In Côte d'Ivoire the CIDT maintains a very dense network of depots, purchasing cotton at every village (DCGT 1988). The costs of transport to the farmer are therefore very small, no more than CFAF5 per kilogram in the mid-1980s (Beenhakker and Bruzelius 1985).

What are the implications of panterritorial pricing for the well-being of farmers and for the revenues of governments? To assess the transport pricing policy embodied in panterritorial pricing requires the simulation of alternative transport policies using equations 1 to 5. This, in turn, requires an assumption about the functional form of the zonal supply function, $S(p)$. There is no information available on the supply function for cotton in Côte d'Ivoire; indeed panterritorial pricing means there is no regional price variation with which to infer response, thus leaving only rather limited time series information. I therefore adopt the form

$$S(p) = p^{\alpha} \tag{6}$$

which has a constant price elasticity of supply denoted by α, and I present a sensitivity analysis for various values of α.

Table 2. *Price Variables in the Cotton Sector in Côte d'Ivoire, 1982/83*
(CFA francs per kilogram)

Variable	*Value*
Export price (p^*)	656
Price received by farmers (p_i^o)	191
Nontransport costs paid by parastatal (c^*)	306
Cost of transport from zone to ginnery per kilogram-kilometer (a^*)	0.067
Cost of transport from ginnery to port per kilogram-kilometer	0.026

Note: The export price is the international price. Nontransport costs are net of all taxes and of the price paid to farmers. All variables refer to one kilogram of ginned cotton. The price received by farmers is constant for all zones because the CIDT pursued panterritorial pricing in 1982/83.

Source: Beenhakker and Bruzelius (1985), especially table 8.

Using the constant-elasticity-of-supply function, the CIDT's implementation of panterritorial pricing (p_i a constant, as given in table 2) is compared with the most desirable pricing scheme (set of p_i's). This latter policy maximizes the well-being of farmers (Π) subject to the constraint of raising an amount of revenue (R^0) for the government equal to that raised under panterritorial pricing. The solution to this transport pricing problem is referred to as the optimal policy. Actually, a tax on land that varies with location is first best, because it does not induce any change in farmers' behavior and therefore dominates even the pricing policy that I term optimal. Such a tax is rarely used, however, and appears to require a degree of administrative capability that is not available.

For any zonal supply function, the optimal policy is found by substituting from equations 1 to 5 into

$$\mathcal{L} = \Pi + \mu(R - R^0) \tag{7}$$

where μ is a Lagrange multiplier and then setting the derivatives of equation 7 with respect to the prices received by farmers to zero. The derivatives are:

$$\frac{d\mathcal{L}}{dp_i} = \frac{\gamma_i S(p_i)}{S(p_i^0)}\left[1 - \mu\left(1 - \frac{\epsilon(p^* - c^* - p_i - a^* t_i)}{p_i}\right)\right] = 0, \quad i = 1, \ldots N, \tag{8}$$

where the elasticity of supply is

$$\epsilon \equiv \frac{dS(p_i)}{dp_i}\,\frac{p_i}{S(p_i)}. \tag{8a}$$

Equation 8 holds for each of the N zones, so there are N equations. Along with the revenue constraint that $R = R^0$, these N equations determine the spatial pattern of zonal prices and the Lagrange multiplier or (shadow) cost of having to raise one additional unit of revenue measured in lost producer surplus, μ. The shadow cost is the only variable that is not specific to a particular zone in the equation for that zone.

In the special case of a constant elasticity of supply, the spatial pattern of prices takes a very simple form and has an intuitive interpretation of an (implicit) optimal subsidy to transport. Newbery (1990) makes a similar point for the subsidization of inputs when output is taxed. As indicated by equation 8, when the supply function has a constant elasticity, the optimal policy is defined by a pair of constants, p_λ and a_λ, such that

$$p_\lambda = \lambda(p^* - c^*) \tag{9a}$$

$$a_\lambda = \lambda a^* \tag{9b}$$

$$p_i = p_\lambda - a_\lambda t_i \tag{9c}$$

and

$$R = R^0. \tag{9d}$$

Table 3. *Alternative Constant-Revenue Pricing Policies for Côte d'Ivoire*

	Transport cost (a^*)					
	CFAF 0.0670 per kilogram			CFAF 0.0335 per kilogram		
	Panterritorial pricing (1)	*Optimal pricing (2)*	*Full-cost pricing of transport (3)*	*Panterritorial pricing (4)*	*Optimal pricing (5)*	*Full-cost pricing of transport (6)*
Elasticity of supply = 0.33						
Government revenue (R)	13.31928	13.31927	13.31927	13.31928	13.31927	13.31928
Producer surplus (Π)	15.00087	15.01116	15.00630	17.13862	17.14105	17.14019
Producer price at port (p)	191	210	223	211	221	227
Implicit cost of transport to producers (a)	0	0.040	0.067	0	0.021	0.034
Elasticity of supply = 0.67						
Government revenue (R)	13.31928	13.31927	13.31927	13.31928	13.31927	13.31928
Producer surplus (Π)	12.00070	12.02921	12.01615	14.89461	14.90100	14.89914
Producer price at port (p)	191	210	223	217	228	233
Implicit cost of transport to producers (a)	0	0.040	0.067	0	0.022	0.034
Elasticity of supply = 1.00						
Government revenue (R)	13.31928	13.31928	13.31928	13.31928	13.31928	13.31927
Producer surplus (Π)	10.00058	10.07077	10.03986	14.14348	14.15609	14.15332
Producer price at port (p)	191	211	223	227	238	243
Implicit cost of transport to producers (a)	0	0.040	0.067	0	0.023	0.034
Elasticity of supply = 1.33						
Government revenue (R)	13.31928	13.31928	13.31928	13.31928	13.31927	13.31926
Producer surplus (Π)	8.57193	8.82360	8.72427	14.69112	14.71218	14.70906
Producer price at port (p)	191	213	224	241	252	257
Implicit cost of transport to producers (a)	0	0.041	0.067	0	0.024	0.034

Note: Government revenue and producer surplus are in billions of CFA francs; producer price and cost of transport are in CFA francs per kilogram. Calculations are for the 53 CIDT zones in table 1. The zonal prices are given by $p_t = p - at_t$.

Source: Author's calculations based on data from CIDT (various years c), Michelin (1989), DCGT (1986), unpublished data provided by the Caisse Central in Abidjan, and Beenhakker and Bruzelius (1985).

Note that p_λ and a_λ differ from p^* and a^* by the same proportion, λ, and it is easy to show that raising a positive amount of revenue, $R^0 > 0$, implies that $p_\lambda < p^*$, that is, $\lambda < 1$. Thus $1 - \lambda$ is both the tax rate on the export of the product and the subsidy rate on transport relative to a situation in which farmers receive the full export price but pay the full cost of transport.

Once the government is raising some revenue, therefore, it is no longer optimal to pass the full cost of transport, a^*, on to the farmer. If the elasticity of supply is constant, transport should be subsidized to the same degree that the export is taxed. In this particular case, the optimal solution has all farmers paying the same tax, $1 - \lambda$, as a proportion of their pretax farmgate price ($p_i = p^* - a^*t_i$). All farmers reduce their production and producer surplus by the same proportion. Therefore the cost of transport must be subsidized so that the percentage tax on the farmgate price is constant regardless of the farmer's location. The linear equation 9c together with equations 9a and 9b does exactly this.[1]

The values of the variables from tables 1 and 2 and equations 1 to 5 as well as various assumptions about the elasticity of supply, α, in equation 6 yield values for revenues, R, and producer surplus, Π, under panterritorial pricing as given in column 1 of table 3. In addition equations 9a to 9d yield the value of the tax rate, $1 - \lambda$, that produces the same revenue as under panterritorial pricing, as given in column 2 of table 3.[2] With such a λ, the value of the producer surplus can be calculated under optimal pricing. For instance, when the elasticity of supply is 0.67, the value of the producer surplus under optimal pricing is 1.0024 (12.02921/12.00070) times the value of the producer surplus under panterritorial pricing (columns 1 and 2 in table 3). When the elasticity of supply is 1.33, which is probably a high value, the corresponding ratio of producer surpluses is 1.0294. Thus, whether the CIDT pursues an optimal policy or one of panterritorial pricing makes relatively little difference to the value of the aggregate producer surplus. Panterritorial pricing is, however, relatively less desirable the more elastic the supply is, that is, the more producers have scope for changing the volume of their output in response to the deviation between panterritorial pricing and the optimal spatial pattern of prices.

In contrast to these results on overall efficiency, switching from a policy of panterritorial pricing to an optimal one has a big effect on the distribution of

1. If the supply function does not exhibit a constant elasticity, then the relationship between the optimal p_i and t_i is nonlinear and could even embody an implicit tax, rather than a subsidy, on transport (see Gersovitz 1989 for a detailed, theoretical discussion of this issue). However, the prospects are slim for econometrically detecting a supply function that deviates significantly from constant elasticity with available data, so, in practice, the question is moot.

2. The value of λ is found by solving numerically the (nonlinear) equation:

$$\Sigma[\gamma_i(p^* - c^* - a^*t_i)^{\alpha+1}](1 - \lambda)\lambda^\alpha - p_0^{\alpha} R^0 = 0$$

where, for each block of tables 3 and 4, R^0 is the value of R in column 1.

Table 4. *Revenue-Maximizing Pricing Policies for Côte d'Ivoire*

	Transport cost (a^*)			
	CFAF 0.067 per kilogram		CFAF 0.0335 per kilogram	
	Panterritorial pricing (1)	Optimal pricing (2)	Panterritorial pricing (3)	Optimal pricing (4)
Elasticity of supply = 0.33				
Government revenue (R)	18.66253	18.66253	18.66253	18.66252
Producer surplus (Π)	4.66563	4.90670	9.26374	9.26713
Producer price at port (p)	80.	91.	133.	139.
Implicit cost of transport to producers (a)	0	0.017	0	0.013
Elasticity of supply = 0.67				
Government revenue (R)	15.25243	15.25242	15.25242	15.25241
Producer surplus (Π)	6.10097	6.49666	11.44762	11.45554
Producer price at port (p)	127.	145.	186.	195.
Implicit cost of transport to producers (a)	0	0.028	0	0.019
Elasticity of supply = 1.00				
Government revenue (R)	13.87734	13.87734	13.87732	13.87731
Producer surplus (Π)	6.93867	7.48075	13.02481	13.03852
Producer price at port (p)	159.	182.	218.	228.
Implicit cost of transport to producers (a)	0	0.035	0	0.022
Elasticity of supply = 1.33				
Government revenue (R)	13.37285	13.37284	13.37282	13.37281
Producer surplus (Π)	7.64163	8.34101	14.57933	14.60056
Producer price at port (p)	182.	207.	240.	251.
Implicit cost of transport to producers (a)	0	0.040	0	0.024

Note: Government revenue and producer surplus are in billions of CFA francs; producer price and cost of transport are in CFA francs per kilogram. Calculations are for the 53 CIDT zones in table 1. The zonal prices are given by $p_r = p - at_r$.

Source: Author's calculations based on data from CIDT (various years c), Michelin (1989), DCGT (1986), unpublished data provided by the Caisse Central in Abidjan, and Beenhakker and Bruzelius (1985).

producer surplus among zones. Column 5 in table 1 gives the ratio of t producer surplus under optimal pricing to that under panterritorial pricing zone for an elasticity of supply of 0.67. With optimal pricing, the most remo zone, number 25, receives only 0.78 of the producer surplus it received un panterritorial pricing, whereas the nearest-in zone, number 51, receives 1. times as much. The coefficient of variation of the proportional gain is 8 percent.

I do not, however, know how these changes in zonal producer surplu would translate into changes in the distribution of farmers' incomes, which is important welfare issue. It all depends on who owns what land and how mu of it. For instance, if all farmers owned an equal share in land everywhere change in the distribution of producer surplus among the zones would not aff the relative well-being of different farmers. In Africa, of course, the expectati is that small farmers depend entirely on agricultural land in one vicinity, changes in the distribution of producer surplus at the zonal level would aff different farmers differently.

The dependence of farmers on land in one area raises further questions: farmers in more remote regions have more or better land—or other sources income—so they are at least as well off as farmers nearer in? In the CIDT zone would not be surprising if cotton were important in the incomes of these farm and if poorer farmers lived in more remote areas. If this is so, the distribution income could become considerably more dispersed by adopting an optimal pr ing policy, while the gains in the aggregate producer surplus would be small farmers' incomes are tied to land within zones rather than widely diversif across zones, considerable redistribution relative to the net gain would have be engineered among farmers in different zones—and in a nondistorting w Otherwise the movement from panterritorial pricing to optimal pricing wou not be a (pareto) improvement that makes no farmer worse off.

If, by contrast, the CIDT were raising more revenue than it currently does estimated in table 3), the situation would be somewhat different with respect efficiency. For instance the price paid to farmers (p^0) that maximizes the gove ment revenue that can be raised under panterritorial pricing is given by:

$$(10) \qquad p^0_{max} = [\Sigma\gamma_i\alpha(p^* - c^* - a^*t_i)]/[\Sigma\gamma_i(1 + \alpha)]$$

as can be derived by setting the derivative of aggregate revenues, equation 4, zero.

Columns 1 and 2 in table 4 give the results of the model when the governm maximizes the amount of revenue that it can raise under panterritorial pricir with transport costs of CFAF0.067 per kilogram-kilometer. For an elasticity supply of 0.67 the revenue-maximizing value of the panterritorial price CFAF127 (table 4, column 1) or only 0.36 times the world price net of costs tl are independent of zonal location (CFAF350 from table 2). In this case an equ amount of revenue can be raised with an optimal pricing policy that has λ 0.418 in equations 9a to 9c and a level of producer surplus 1.0

(6.49666/6.10097) times that obtained with panterritorial pricing—a moderate difference. If the elasticity of supply is 1.33, the ratio of producer surplus under optimal pricing to that under panterritorial pricing is 1.092. Movement from the current situation (table 3, column 1) to revenue maximization (table 4, column 1) brings a small increase in revenue (about CFAF54 million) but a large decrease in producer surplus (about CFAF930 million). There is almost a 10 percent potential gain (about CFAF699 million) in producer surplus in moving from panterritorial pricing to optimal pricing (table 4, columns 1 and 2). In general, as the government raises more revenue, the deadweight loss of panterritorial pricing rises relative to that occurring under the optimal policy. Furthermore it is simply impossible to raise more revenue than that given in column 1 of table 4 if the constraint of panterritorial pricing is maintained, although with an optimal policy it would be possible to raise more if desired.

In Côte d'Ivoire cotton is produced in a geographically compact area with good transport compared with many other export hinterlands in Africa, yet the government's choice of an (implicit) transport pricing policy can affect producer surplus. For it to matter significantly, however, the government would have to be raising more revenue than seems to have been its practice as estimated in table 3. For example, in table 4, columns 1 and 2, more revenue is raised than in table 3, and when $\alpha = 1.33$ distortions are estimated to be moderately large. By contrast, if the hinterland were more dispersed or if transportation per kilometer were more expensive, the costs of panterritorial pricing relative to optimal pricing would be higher. In simulations corresponding to those reported in tables 3 and 4, the cost of transport, a^*, was doubled to 0.134. Even for this value, net returns to exporting cotton were greater than the cost of transport for all zones ($p^* - c^* > a^* t_i$). The result, corresponding to table 4, columns 1 and 2, for $\alpha = 1.33$ but for $a^* = 0.134$ was a producer surplus ratio of 1.21 in favor of optimal pricing as opposed to the much lower factor of 1.092 reported in table 4 for $a^* = 0.067$.

Tables 3 and 4 provide information to assess the benefits from investments in transport infrastructure as represented by a decrease in the cost per kilogram-kilometer, a^*. Revenue amounts shown in columns 4 and 5 of table 3 are the same as in columns 1 and 2, respectively, but transport costs are only half as much, and the same is true in table 4, when columns 3 and 4 are compared with columns 1 and 2, respectively.[3] The values in table 3 show that the gains in

3. Column 4 of table 3 is derived by finding numerically the p_1 that solves the (nonlinear) equation:

$$\Sigma[\gamma_i (p^* - c^* - a^* t_i)] p_1^{\alpha} - \Sigma \gamma_i p_1^{\alpha+1} - p_0^{\alpha} R^0 = 0$$

where p_0 is the value of p in column 1. Column 5 is determined from column 4 in the same way that column 2 is determined from column 1 (see footnote 2). The same procedure is followed to compute columns 3 and 4 of table 4.

producer surplus from a transport improvement under panterritorial pricing are larger than under optimal pricing.[4] Nonetheless the magnitude of the differences is small. For $\alpha = 1.33$, the most extreme case in table 3, the gain under panterritorial pricing is CFAF6.12 billion, compared with CFAF5.89 billion under optimal pricing. These results carry over qualitatively to table 4, although raising more revenue accentuates the difference between the gains under the two pricing rules. For an elasticity of 1.33 the gain from halving transport costs under panterritorial pricing is CFAF6.94 billion; it is CFAF6.26 billion under optimal pricing. In this hypothetical example, whether a change in pricing policy is adopted may well determine whether an infrastructural investment is justified.

Investment in transport under panterritorial pricing yields greater benefits than under optimal pricing, because production is more dispersed under panterritorial pricing. Decisions about transport investments and pricing are interdependent. Of course the producer surplus is still highest under optimal pricing for a given level of infrastructure. Panterritorial pricing should not be adopted only to get higher apparent gains from transport investments. Rather, the analysis indicates that a move from panterritorial pricing to optimal pricing, if feasible, is an alternative way to realize some of the gains from a transport investment under panterritorial pricing without the corresponding costs. Thus the change in pricing may obviate the need for investment in transport.

The discussion comparing panterritorial pricing and optimal pricing generalizes easily to any other pricing scheme. For instance an export tax (at rate τ) with private (full-cost) transportation would imply that the price received by farmers is given by

$$p_i = (1 - \tau)(p^* - c^*) - a^* t_i. \tag{11}$$

Such a scheme has superficial appeal because it may be thought to embody a user charge for transport, which is desirable in many other situations. From equations 9a to 9d, however, such a scheme is suboptimal if the government is raising some revenue ($R > 0$, $\lambda > 0$, and $\tau < 1$), as in the case of cotton in Côte d'Ivoire. As noted, the reason is that charging the full cost of transport means that a tax at the port results in higher percentage taxation of the farmgate price as distance rises.

The results of full-cost pricing of transportation are illustrated in columns 3 and 6 of table 3.[5] In terms of producer surplus, full-cost pricing of transport is better than panterritorial pricing, generally making up about 60 percent of the

4. That is, the difference in Π between columns 1 and 4 is larger than the difference in Π between columns 2 and 5.

5. The effect on the producer surplus associated with a given government revenue can be found by solving for the value of τ that produces a revenue equal to R^0 from equations 1, 2, and 4 and then substituting the consequent set of p_i into equations 3 and 5.

gap between panterritorial pricing and optimal pricing. Moving from p
ritorial pricing to full-cost pricing causes a larger dispersion of zonal gains
losses than does moving from panterritorial to optimal pricing. For the para
ters of the example given in table 1, column 5, the summary statistics for
ratio of producer surplus under full-cost pricing to that under panterrito
pricing are: mean, 0.98; standard deviation, 0.12; minimum, 0.64; and ma
mum, 1.18. Finally, because full-cost pricing results in less dispersed produ
tion, it makes investment in transportation infrastructure less attractive tha
does either alternative (compare columns 1 to 3 with columns 4 to 6 of table 3).

III. The Density of Depots and Backhauling

Farm outputs are often gathered at depots before transport to processing plants or to the point of export. The number and location of these depots can have significant economic effects. One important attribute of a depot system is the extent of backhauling that it engenders. Here I bend the conventional definition of backhauling to mean the movement by farmers of their output *away* from the point of processing, export, or ultimate consumption. Backhauling can increase the transport costs of the sector unnecessarily.

Not all marketing systems engender backhauling. For instance, if depots exist in every village and the marketing authority adopts panterritorial pricing, then a farmer gets the same price for output no matter where it is sold. Therefore the farmer sells the output at the nearest depot, which is very close to where the crop is harvested, and there is (almost) no backhauling. This situation prevails in Côte d'Ivoire, where the CIDT operates one depot in each cotton-growing village in the sector.

Another situation in which there need not be an incentive to backhaul occurs if the marketing parastatal equates the price at any depot to the price received at the final point of sale less the full cost of transport to that point. If, in addition, the private sector has the same (or lower) costs of transport to the depot, it never pays to send produce away from the destination to which the parastatal wants it to go. The depots nearer the ultimate destination pay a sufficiently higher price so that farmers send their output to those depots even if they are farther away than other depots.

The preceding section of this article, however, establishes that full-cost pricing of transport is not optimal *if* the government is raising revenue from the sector. In particular, when the elasticity of supply is constant, equations 9a to 9d show that the optimal spatial pattern of prices embodies a subsidy to transport. In this case it is as if the private sector has more costly transport than the parastatal. It may become privately profitable to move output to a depot that is nearer to the farm but farther from the destination to which output is being moved by the parastatal. Backhauling then occurs. Although depots nearer to the parastatal's ultimate destination pay higher prices, they are insufficiently higher to guarantee

that it is never profitable to ship to depots that are farther from the ultimate destination but very much nearer to the farmer.

If there are very many depots, then the amount of backhauling is unimportant, because the depot that is farther from the ultimate destination but nearer to the farmer and the depot that is farther from the farmer but nearer to the ultimate destination are both absolutely very near the farmer. If depots are expensive to set up, however, there will be of necessity few of them, and then backhauling can be a problem. Backhauling increases the deadweight loss from raising a given amount of revenue and lowers the maximum amount of revenue that the government can raise (Gersovitz 1989).

Panterritorial pricing can further worsen the problem of backhauling when there are few depots because, by equalizing all depot prices, it gives no incentive for farmers to sell to depots nearer the ultimate destination instead of the depot that is nearest to the farmer. How serious the backhauling problem is depends not just on the number of depots but on their location, so little of generality can be said. Nonetheless some examples suggest what to look for and what to expect from different schemes that may be met in practice.

If there are few depots all with the same producer price, but all are clustered near the ultimate destination, then backhauling is likely to be a small problem. For example, the government may adopt the principle of paying only one price but, in practice, buy output only at the ultimate destination. Thus the pricing scheme is one of full-cost pricing of transport because, although the parastatal pays only one price, it pays this price at only the ultimate location. In this case a second depot, paying the same price as the first and located (optimally) quite near it, can decrease the deadweight loss despite inducing some backhauling. This occurs because, as noted, full-cost pricing of transport is not optimal. A second depot located near the first and paying the same price mimics the transport subsidy implicit in equations 9a to 9d that minimizes deadweight loss and at the same time engenders little backhauling. The net gain from the extra depot is, however, relatively small.[6]

By contrast, when the placing of depots is constrained and there are few of them relative to the distance to the ultimate destination, severe losses can occur from backhauling. For instance, if production occurs at one ton per kilometer along a road of length D (as measured from the ultimate destination), then, with only one depot at the ultimate destination, the total ton-kilometers to transport the whole crop to the ultimate destination is proportional to $D^2/2$. By contrast, if a second depot, paying the same price, is located D from the ultimate destination (the absolutely worst place for it along the road), then all output of amount $D/2$ produced along the road from point $D/2$ to D is nearer to the second depot

6. The model of equations 21 and 22 of section VI of Gersovitz (1989) can be used to prove this statement by imposing the additional constraint (in the notation of Gersovitz 1989) that $p_0 = p_1$ and then maximizing with respect only to x_D and the common value of p_0 and p_1.

and is brought there ($D^2/8$ ton-kilometers in total).[7] *This output of amou D/2 produced from D/2 to D then has to be brought back a distance of D to th ultimate destination ($D^2/2$ ton-kilometers). Total transport costs are proportional to $3D^2/4$, including the ($D^2/8$) ton-kilometers generated by the production between point D/2 and the ultimate destination at point 0. Transport costs rise by 50 percent compared with purchase at only the ultimate destination.* Thus a combination of panterritorial pricing, few depots, and bad locations can induce large excess transport costs. In general, if production is uniformly distributed along a road with many (n) depots, each collecting output from a radius of r, the amount of backhauling is proportional to nr^2, providing a rough guide to the extent of backhauling.

IV. Conclusions

This article provides diagnostic tests for analyzing the transport policies that are explicitly or implicitly pursued through state marketing in Africa. A central result of the analysis is the interrelationship among rules for the taxation of exports, the pricing of transport, and investment in transport. The models are fairly general and can be applied to readily available data.

If the government uses an export tax to raise revenue, it may be optimal to subsidize transport to lessen the loss in total producer surplus from raising a specified amount of revenue. In the particular case of a supply function with constant elasticity, transport should be subsidized at the same rate as the export is taxed. This result has direct relevance to the debate on the pricing of diesel fuel and kerosene. The usual presumption is that, while it is desirable to subsidize kerosene because the poor use it as a cooking fuel, such a subsidy would result in unacceptable distortions because kerosene can be substituted for diesel as fuel in the transport sector (Newbery and others 1988). This article suggests that some subsidization of transport may be desirable, thus moderating the conflict between kerosene and diesel pricing policies.

The optimal policy can be contrasted with two suboptimal policies: panterritorial pricing and full-cost pricing of transport with export taxation. For cotton in Côte d'Ivoire, the cost of the current policy of panterritorial pricing does not seem to be large relative to the optimal policy. Full-cost pricing closes a little more than half the gap between panterritorial pricing and optimal pricing. Elsewhere in Africa, however, the agricultural hinterland is much larger, production more dispersed, and the gains from moving to optimal pricing are likely to be correspondingly larger. Furthermore, if the network of depots at which output is bought is badly designed, it can lead to large losses, especially when

7. As in footnote 6, the model of equations 21 and 22 of section VI of Gersovitz (1989) can be used to prove this statement by imposing the additional constraint that $p_0 = p_1$ and looking at arbitrary values of x_D and the common value of p_0 and p_1.

panterritorial pricing is adopted. Such is not the case in Côte d'Ivoire, if the current dense network of depots is not costly to operate and maintain. Nonetheless, if the government of Côte d'Ivoire were ever to seek more revenue from the cotton sector, the losses from staying with panterritorial pricing would increase. The conclusions on the harmful effects of panterritorial pricing depend critically on how much revenue the government raises from the agricultural sector.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Beenhakker, Henri L., and Nils Bruzelius. 1985. "Transport and Marketing of Agricultural Products in the Ivory Coast." World Bank, Infrastructure and Urban Development Department, Washington, D.C. Processed.

CIDT (Compagnie Ivoirienne pour le Developpement des Textiles). Various years a. *Rapport Annuel d'Activité*. Bouaké, Côte d'Ivoire.

———. Various years b. *Rapport Annuel (Synthese)*. Bouaké, Côte d'Ivoire.

———. Various years c. *Rapport Annuel (Annexes Statistiques)*. Bouaké, Côte d'Ivoire.

DCGT (Direction et Contrôle des Grands Travaux). 1986. *Plan National de Transport: Coton*. Abidjan, Côte d'Ivoire.

———. 1988. *Plan National de Transport: Coton*. Abidjan, Côte d'Ivoire.

Gersovitz, Mark. 1989. "Transportation, State Marketing, and the Taxation of the Agricultural Hinterland." *Journal of Political Economy* 97 (5):1113–37.

Michelin. 1989. "Carte Routiere et Touristique, Côte d'Ivoire." Paris.

Newbery, David M. G.. 1990. "Optimal Trade Taxes on Agriculture in Developing Countries." *Economic Journal* 100:180–92.

Newbery, David M., Gordon A. Hughes, William D. O. Paterson, and Esra Bennathan. 1988. *Road Transport Taxation in Developing Countries: The Design of User Charges and Taxes for Tunisia*. World Bank Discussion Paper 26. Washington, D.C.

Walters, Alan A. 1968. *The Economics of Road User Charges*. Baltimore, Md.: The Johns Hopkins University Press.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO 2: 233-254

The Economics of Farm Fragmentation: Evidence from Ghana and Rwanda

Benoit Blarel, Peter Hazell, Frank Place, and John Quiggin

Farm fragmentation, in which a household operates more than one separate parcel of land, is a common phenomenon in Sub-Saharan Africa. Concerned by the perceived costs of fragmented as opposed to consolidated holdings, several countries have implemented land consolidation programs. But these interventions overlook the benefits that land fragmentation can offer farmers in managing risk, in overcoming seasonal labor bottlenecks, and in better matching soil types with necessary food crops. This article uses household data from Ghana and Rwanda to discuss the incidence and causes of fragmentation. It then formally tests the relation between fragmentation and land productivity and risk reduction. The conclusion is that consolidation programs are unlikely to lead to significant increases in land productivity and may actually make farmers worse off. Policymakers should focus instead on reducing the root causes of fragmentation: inefficiencies in land, labor, credit, and food markets.

The existence of fragmented landholdings is an important feature in less-developed agricultural systems. The costs of fragmentation include increased traveling time between fields (hence lower labor productivity and higher transport costs for inputs and outputs), negative externalities (such as reduced scope for irrigation and soil conserving investments as well as the loss of land for boundaries and access routes), and greater potential for disputes between neighbors. In light of these costs numerous land reform policies have been aimed at enforcing, or at least subsidizing, the consolidation of holdings. These policies are premised on the assumption that fragmentation is necessarily inefficient and that agricultural production and social welfare can be increased through land consolidation.

Such policies have already been tried in some African countries (for example, Kenya and Tanzania) and are now being considered by others (for example, Rwanda). Yet little is really known about the incidence of farm fragmentation in Sub-Saharan Africa by region and farm type and about its origins, causes, and

Benoit Blarel is with the Africa—South Central and Indian Ocean Department and Peter Hazell is with the Agriculture and Rural Development Department, both at the World Bank; Frank Place is with the International Council for Research in Agroforestry in Nairobi, Kenya; and John Quiggin is with the Australian National University in Canberra.

impact on land productivity. Although there are negative aspects to farm fra mentation, there are also reasons why it may be beneficial to farmers (such a reducing risk, easing seasonal labor bottlenecks, and enhancing household-leve food security). In fact some degree of farm fragmentation may be desirable and might best be viewed as a rational response by farmers to the economic and institutional environment in which they live. If so, attempts to consolidate holdings beyond some optimal degree could actually reduce farm productivity, and interventions should be focused instead on those features of the rural economy that create the need for fragmentation in the first place.

This article examines the circumstances under which farm fragmentation may be economic to farmers, especially in Sub-Saharan Africa, and relates these circumstances to prevailing conditions in selected regions in Ghana and Rwanda. Available survey data are then used to test the relations among farm fragmentation, and land productivity and risk management.

I. Explanations of Fragmentation

The explanations of land fragmentation offered in the literature fall into two broad categories (McPherson 1982; Bentley 1987). The first treats fragmentation as an exogenous imposition on farmers and consists of what we shall call "supply-side" explanations. The second views fragmentation as primarily a choice variable for farmers and can therefore be described as "demand-side" explanations. Supply-side explanations invariably conclude that fragmentation has adverse effects on agricultural production. Demand-side explanations presume that farmers will, given free choice, choose levels of fragmentation that are beneficial. Imperfections in factor or commodity markets play a key role in both types of arguments.

Supply-Side Explanations

Several forces have been widely cited as causing or contributing to involuntary fragmentation. The most frequently cited are partible inheritance and population pressure resulting in land scarcity (Anthony and others 1979; Binns 1950; Holmberg and Dobyns 1969; Hyodo 1963; Leibenstein 1957; Thirsk 1964; Wolf 1966; World Bank 1978). Many authors argue that partible inheritance logically leads to fragmentation when farmers desire to provide each of several heirs with land of similar quality. Likewise extreme land scarcity may lead to fragmentation as farmers in quest of additional land will tend to accept any available parcel of land within reasonable distance of their house (Farmer 1960).

These factors explain why a young farmer might begin with a fragmented holding. However, as McCloskey (1975) points out, they do not explain the persistence of fragmentation in the face of economic incentives for consolidation. Such persistence indicates significant imperfections in the land market. Some economists (for example, Bauer and Yamey 1957; Lipton 1968; Sargent 1952) claim that land markets themselves are highly fragmented, with few will-

ing sellers. Dorner (1977) cites multiple interests over parcels as restricting the potential supply of land, because unanimous agreement to sell is difficult to achieve. McKinnon (1973) stresses incomplete credit markets and the resulting inability of many farmers to finance land acquisitions.

Another supply-side factor is the breakdown of common property systems under the pressure of population growth. This breakdown has led to increased fragmentation in, for example, Kenya (King 1977) and eastern Nigeria (Udo 1965). A number of authors have demonstrated that fragmentation in certain areas is a consequence of egalitarian objectives on the part of the communal authority (Dahlman 1980; Georgescu-Roegen 1969; Grigg 1970; Quiggin 1988). State laws that restrict land transactions also limit possibilities for land consolidation. Finally, nature itself may limit the boundaries of arable parcels (for example, waterways and wastelands) so that expansion of farm size requires the acquisition of separate pieces of land.

The supply-side explanations, while plausible, are not sufficient to explain fragmentation in all the areas in which it is found. First, even where land markets afford farmers opportunities for consolidation, fragmentation persists (see section III). Second, fragmentation has developed in the absence of land scarcity (for example, in areas of Kenya, Zambia, and The Gambia [McPherson 1982]). Third, ancestors continue to bestow heirs with scattered holdings, a practice that would seemingly be halted if fragmentation was largely detrimental (Douglass 1969; Leach 1968). The argument that partible inheritance is designed for equity reasons runs into difficulty when it is observed that subdivision and fragmentation levels are eventually "checked" after reaching specific levels (noted in India by Hopper 1965, in Mexico by Downing 1977, and in Sri Lanka by Leach 1968). These examples suggest that other factors may be important in explaining fragmentation.

Demand-Side Explanations

Demand-side explanations of farm fragmentation presume that the private benefits of fragmentation exceed its private costs. That fragmentation might benefit farmers follows from the realization that land is not homogenous. Parcels differ with respect to soil type, water retention capability, slope, altitude, and agroclimatic location. Recognizing this, Buck (1964) and Johnson and Barlowe (1954) were among the first to note that by operating parcels in different locations, farmers are able to reduce the variance of total output and hence final consumption. This is partly because the scattering of parcels reduces the risk of total loss from flood, drought, fire, and other perils and also because farmers can more efficiently diversify their cropping mixtures across different growing conditions. Other risk-spreading mechanisms, such as insurance, storage, or credit, also reduce variations in household consumption. Therefore fragmentation for risk reduction should persist only if these alternatives are either not available or are more costly (Carter and Matlon nd; Charlesworth 1983; Fenoaltea 1976 ; Hyodo 1963; Ilbery 1984; McCloskey 1976; Thompson 1963).

Table 1. *Characteristics of Regions Studied for the Economics of Fragmentation, 1987–88.*

Region	*Location*	*Population density (persons/square kilometer)*	*Sampled households*	*Households with non-farm income (percent)*	*Median farm size (hectares)*	*Median number of parcels*	*Mean value of Simpson index*	*Median household size*
Ghana								
Anloga	Southeast	384	115	89.6	0.26	5	0.64	7
Wassa	Southwest	30	150	69.3	13.24	5	0.62	7
Ejura	Central	20	158	30.4	2.57	1	0.23	6
Rwanda								
Ruhengeri	Northcentral	367	72	47.2	0.70	7	0.69	5
Butare	Southcentral	342	80	25.0	0.91	7	0.62	6
Gitarama	Central	279	80	25.0	0.83	5	0.54	6

Source: Authors' calculations based on World Bank survey data.

Another explanation for fragmentation was developed by Fenoaltea (197€ for medieval England. He argued that because of transaction costs in labc markets, the scattering of parcels enabled farmers to better fulfill their season labor requirements and consequently to obtain higher yields. If the labor marke does not work at all, labor supply is fixed by household size, and the need fc temporally spreading labor requirements is great. Even if labor markets exis the costs of supervision may induce farmers to scatter parcels and supervise small number of workers at a time, rather than watch over a large number c hired workers on a consolidated holding at peak periods. This approach is mo effective when different types of land are suitable for different crops (henc when fragmentation facilitates diversification) or when different parcels of lan offer sufficient diversity in climatic conditions that the same crop can be sta gered over a wider range of planting dates.

Commodity market failures may also cause fragmentation to have a positiv impact on productivity. When such failures occur, a subsistence mode may b adopted in which several products are raised for household consumption, rath than purchased with the proceeds of cash crop sales. If different land types c ecozones are suitable for cultivating different crops, then the required diversit can best be obtained from a fragmented landholding (Netting 1972). Becau this argument assumes that all of the products required can be produced by single household, and hence within a single village, trade within the village mu also be more costly than land fragmentation. This seems most likely to happe when there is considerable uncertainty about relative price movements, esp cially for important foods. Farmers will not want to specialize in less essenti foods if there is a real risk that they cannot procure sufficient amounts c essential foods through trade. These price risks can be particularly real in Afric given that the risk-pooling effects of interregional trade are hindered by poc infrastructure and marketing systems.

Farmers might also want fragmented holdings if, holding farm size constan there are diseconomies of scale with respect to the size of individual parcel Where this phenomenon occurs, however, it probably reflects malfunctionin labor markets; farmers are unable to procure adequate labor to meet season peaks in the requirements for large parcels.

II. Evidence of Fragmentation

Data on fragmentation were obtained from three regions in Ghana an Rwanda as part of a larger study of land rights in Sub-Saharan Africa (Migo Adholla and others 1991). Nearly 150 households were surveyed in each of th Ghanaian regions and about 80 in each of the Rwandan areas (table 1). Th data were collected between 1987 and 1988 and included characteristics of th household (such as age, education, and occupation of all individuals; number c regular workers; nonfarm income; wealth; and use of credit), characteristics c the household head (such as farming experience, place of birth, and local offic

held), characteristics of the farm (such as size and number of parcels), characteristics of each parcel (such as mode of acquisition, soil fertility, dista from house, topography, size, rights of transfer and use, land improveme made, crops grown, and inputs and output in the past season).

In measuring fragmentation, we primarily use the Simpson index, defined $1 - \Sigma_i A_i^2/A^2$, where A_i is the area of the *i*th plot and $A = \Sigma_i A_i$ is the total farm area. A value of zero indicates complete land consolidation (one parcel only), while the value of one is approached by holdings of numerous parcels of equal size. However, because the Simpson index is sensitive to dispersion in the size of the parcels as well as to their number, we also use the number of parcels as an alternative fragmentation measure. Values for the Simpson index and number of parcels are for individual farms. To look at farm fragmentation in a region, a frequency distribution is constructed of the percentage of farms at different levels of the indicators.

Ghana

The regions in Ghana for which we have data include Anloga, a coastal strip in the southeast; Ejura, a savanna area in the center; and Wassa, a forest region in the southwest. The Anloga region has a very high population density (384 persons per square kilometer), which is unusually high for Ghana. The other two regions are much more land-abundant, with densities of around 25. Subtle differences in topography or altitude within the three regions can have a measurable impact on cropping strategies. For instance the Anloga region has the least ecological diversity, yet location, such as proximity to lagoons, is important because it is related to the risk of flooding. The Anloga area is predominantly a shallot-growing area, with only minor areas devoted to other crops. Cocoa predominates in the Wassa area, even though the land is only marginally suitable for cocoa. In Ejura farmers produce a variety of food crops along with groundnuts, and, unlike all the other study regions, many cultural operations are tractorized.

The evidence reported in table 2 indicates that farm holdings in Anloga are highly fragmented. The median number of parcels is 5.0, and the median value of the Simpson index is 0.7. The parcels are, however, generally located close enough to the house to be readily serviced by the household; the average distance is 1.2 kilometers. Greater fragmentation does not result in greater average distances for farmers; the correlation between the Simpson index and the average distance to a parcel is −0.091 (significant at only the 17 percent confidence level), indicating that, if anything, parcels on more fragmented farms are generally closer to the homestead.

For the Wassa region the median number of parcels per farm is 5.0, and the median value of the Simpson index is 0.66. Farmers must travel 1.8 kilometers on average to reach their parcels. Again the average distance to parcels (across households) is not significantly related to the Simpson index of fragmentation (the correlation of 0.059 is only significant at a level of 0.24). The average

distance to parcels is lower for migrant farmers than for indigenous farmers (1.3 versus 2.3 kilometers, respectively) because migrants typically acquire land before building a dwelling, whereas indigenous farmers are more likely first to become established in an existing dwelling or settlement and then to acquire land. Farms managed by migrants are not significantly different from those managed by indigenous farmers with respect to the number of parcels or the Simpson index.

Fragmentation is not as prevalent in Ejura. Table 2 shows that more than 50 percent of farmers operate only one parcel, while 94.4 percent operate three or fewer parcels. The Simpson index is correspondingly low. For 84 consolidated farms the index value is zero. For multiple-parcel farms the index is below 0.55 in most cases and has a maximum of 0.75. The mean value is 0.23. The distance to parcels is greater in Ejura. More than 20 percent of farmers must travel an average of 4.7 kilometers or more to reach their parcels. This is not a problem specific to fragmented farms; similar distances must be traveled to parcels on consolidated and fragmented holdings. The correlation between average distance and number of parcels is 0.002, implying virtual independence of the two variables.

Table 2. *Fragmentation of Operated Land in Three Regions in Ghana, 1987–88*
(percentage of households)

Measure of dispersion	*Anloga*	*Wassa*	*Ejura*
Simpson index			
0–0.5	20.9	20.7	87.3
0.5–0.7	27.0	43.3	9.5
0.7–0.8	30.4	28.0	3.2
Over 0.8	21.7	8.0	0.0
Mean*	0.64	0.62	0.23
Median*	0.71	0.66	0.00
Number of parcels			
1	2.6	0.0	53.2
2–3	25.2	11.3	41.2
4–5	27.0	43.3	4.4
6–7	22.6	33.3	1.3
8 or more	22.6	12.0	0.0
Mean*	5.67	5.52	1.70
Median*	5.00	5.00	1.00
Average distance to parcel (kilometers)			
0–0.5	13.9	15.3	1.3
0.5–1	32.2	23.3	7.6
1–1.5	25.2	16.0	6.3
1.5–2	15.7	14.0	12.7
Over 2	13.0	31.3	72.2
Mean*	1.21	1.80	3.57
Median*	1.10	1.50	3.00

*Expressed in relevant units, not percentages.
Source: Authors' calculations based on World Bank survey data.

Table 3. *Fragmentation of Operated Land in Three Prefectures in Rwanda, 1987–88*
(percentage of households)

Measure of dispersion	*Pooled*	*Ruhengeri*	*Butare*	*Gitarama*
Simpson index				
0–0.2	7.3	2.8	7.5	11.3
0.2–0.4	10.3	5.6	10.0	15.0
0.4–0.6	21.1	12.5	20.0	30.0
0.6–0.8	35.8	44.4	41.3	22.5
0.8–1.0	25.4	34.7	21.3	21.3
Mean*	0.61	0.69	0.62	0.54
Median*	0.66	0.75	0.66	0.57
Number of parcels				
1–2	9.5	8.3	2.5	17.5
3–4	22.4	19.3	18.8	28.8
5–7	27.6	27.8	30.0	25.0
8–10	22.4	25.0	27.5	15.0
Over 10	18.1	19.4	21.3	13.8
Mean*	7.1	7.7	7.9	5.9
Median*	6.0	7.0	7.0	5.0
Topographical dispersion[a]				
Zero	25.4	38.9	6.3	32.5
0–0.2	22.4	9.7	37.5	18.8
0.2–0.3	12.9	8.3	16.3	13.8
0.3–0.4	11.6	12.5	15.0	7.5
0.4–0.5	14.6	12.5	12.5	17.5
Over 0.5	13.4	18.1	12.5	10.0
Mean*	0.24	0.25	0.26	0.21
Median*	0.22	0.23	0.23	0.17
Average distance (minutes)[b]				
0–3	22.4	11.1	16.3	38.8
3–6	16.8	15.3	21.3	13.8
6–9	20.3	20.8	20.0	20.0
9–15	21.1	30.6	26.3	7.5
Over 15	19.4	22.2	16.3	20.0
Mean*	9.8	10.9	9.6	9.1
Median*	7.5	10.0	7.9	5.4

*Expressed in relevant units, not percentages.

a. An index similar to the Simpson index is used to measure topographical dispersion. The measu $1 - [\Sigma_i t_i^2 / A^2]$, where t_i is the area of land located in topographical location i, and A is the total farm a The index is equal to zero when all land is located in a single topographical location.

b. We use time rather than kilometers because this is a more meaningful measure, given the steep, l terrain.

Source: Authors' calculations based on World Bank survey data.

The Simpson index is positively correlated with farm size in Anloga (0.305) but is unrelated to farm size in Wassa and Ejura. In all three regions, a higher level of fragmentation is associated with lower average parcel sizes.

Rwanda

Rwanda has one of the highest population densities in Sub-Saharan Africa (445 inhabitants per square kilometer of arable land in 1986), as well as one of the highest population growth rates (3.6 percent a year between 1978 and 1986). The ratio of population to total land is high in each of the three study prefectures: Ruhengeri, Butare, and Gitarama. Pressure on existing farmland is made worse by the lack of new land available for agriculture and the lack of off-farm opportunities. The microecological variation of land across Rwanda is tremendous. It is a country dotted with steep hills, where altitudes and slopes change dramatically within shouting distance. A wide variety of crops is grown in each prefecture as well as on most individual farms. The common crops are coffee, sorghum, beans, bananas, sweet potatoes, groundnuts, maize, and Irish potatoes.

The first block of table 3 shows the distribution of the Simpson index for operated parcels by prefecture. Holdings are most fragmented in Ruhengeri and least in Gitarama. Fragmentation declines with farm size; the correlation is statistically significant at -0.201. Gitarama farms typically have fewer parcels than those in Ruhengeri or Butare. This is easily explained by the higher incidence of paysannat farms (settlement schemes farms), which consist of one parcel. Farms are also less dispersed topographically than spatially in all three prefectures. This indicates that farmers operate separate parcels in similar topographical locations, a practice that is most prevalent in Ruhengeri and Gitarama but relatively rare in Butare. The average distance, measured in minutes, between the house and the operated parcels is lowest in Gitarama and highest in Ruhengeri. For the median farm a household member must travel 5.4 minutes to reach a parcel in Gitarama, as opposed to 7.9 minutes in Butare and 10.0 minutes in Ruhengeri.

III. Causes of Fragmentation in Ghana and Rwanda

Based on our review in section I, we would expect fragmentation in areas of low population density to be driven primarily by demand. Demand in these areas, while originating from imperfections in the credit, labor, or food markets, will depend on the extent of soil and agroclimatic diversity within the community and hence on the possibilities for staggering labor tasks and diversifying into different crops. In regions where land is scarce, however, supply-side factors are likely also to come into play, especially where land markets are too limited to permit desired levels of farm consolidation. Other things being equal, we would therefore expect fragmentation to be greater in more highly populated regions.

Our regional data confirm a positive relation between population density and

the Simpson index of farm fragmentation (table 1). Wassa, in Ghana, is the exception: even though it has a low population density similar to that of Ejura, also in Ghana, the level of fragmentation is comparable with regions with more than 10 times its population density. Fragmentation in Wassa, however, is largely determined by the scarcity of good cocoa land. Farmers generally seek level areas to establish cocoa parcels, and these areas tend to be small and separated by numerous hills and waterways.

A positive relation between population density and fragmentation does not in itself prove that fragmentation is predominantly driven by supply. It is also necessary to show that land markets fail. This is difficult to verify, given the available data, but some inferences are possible. Land market transactions account for less than half of all parcel acquisitions in our study regions in Rwanda and less than one-third of those in Ghana. Land purchases are particularly rare—except in Ruhengeri (in Rwanda) and Wassa (in Ghana), where they account for about 18 percent of all parcel acquisitions—and they seem to be unrelated to population density. Conversely, temporary land transfers (renting and borrowing) are more common in the more highly populated regions, although relatively rare in the regions with low population densities. One would expect these temporary transactions to provide considerable scope for farmers to consolidate their operated holdings, even if the possibilities for permanent consolidation of owned land are more limited.

But are temporary land transactions used to consolidate operated holdings, especially in the more densely populated areas? We do not have enough data to explore this question for Ghana. In Rwanda, however, apart from a small percentage of large farms that only lease out (rent or lend) land, all households that lease land actually increase the average fragmentation of their operated holdings and the average distance that they must travel to their plots. This is true even of the 14 percent of farmers who "swap" parcels through simultaneous leasing arrangements. Although one should expect smaller farms to lease any available piece of land in a land-scarce situation, the fact that parcel "swapping" is also used to increase fragmentation suggests there may be significant demand for fragmentation in Rwanda. We cannot therefore conclude that fragmentation is a predominantly supply-driven phenomenon in these highly populated areas.

If farm fragmentation in the highly populated study regions is not purely driven by supply, then we should not necessarily expect it to reduce the average productivity of land. Moreover the effect on productivity will depend on the degree of heterogeneity in soil and agroclimatic conditions, and hence the possibilities for efficient crop diversification or the staggering of labor tasks.

Because Anloga and Wassa, in Ghana, are primarily monocrop regions, the advantages of fragmentation are likely more modest there. Nevertheless in Anloga topographical location, soil fertility, and soil moisture retention vary slightly, and, while in dry periods the depressed areas perform relatively better, in wet periods they are more susceptible to flooding. By operating parcels in areas of low and high flood risk, farmers can reduce the likelihood of a complete

disaster as well as contain the variability in total farm yields. The Wassa region is characterized by rolling hills so that spatial separation of flat parcels may mitigate the risks of severe crop failure caused by pests or disease.

Ejura, however, has a wide range of soil types that facilitate efficient crop diversification. The survey data show that fragmented farms have more diversified cropping patterns than do consolidated farms. Land is also relatively abundant in Ejura, and, with its low population density there is little reason to expect that fragmentation is supply-induced. Given the widespread use of tractors in the region, however, we expect the demand for fragmentation to be muted by likely economies of scale in parcel size.

In Rwanda the extensive soil and agroclimatic diversity suggests that fragmentation would offer considerable scope for diversifying crops and staggering planting dates. These advantages may explain why even those farmers who "swap" land through simultaneous leasing arrangements have farms that become more fragmented. Such arrangements may also significantly cushion any efficiency losses arising from supply-induced fragmentation.

In conclusion, there are a priori grounds for expecting that farm fragmentation is driven by demand as well as supply factors in five of the six study regions. Fragmentation in Ejura, which, as shown in table 2, is not widespread, is likely to be largely a demand-driven phenomenon.

IV. An Econometric Model

The measurement of the effect of fragmentation on land productivity requires the formulation of an appropriate econometric model. There are three types of household decision variables of interest: the level of farm fragmentation, F; the parcel-level use of direct inputs, L_i; and the current stock of land improvements on each parcel, I_i. These variables, together with other shifter variables, determine parcel yield, Y_i, in the following structural model:

$$Y_i = f(L_i, I_i, F, X_{1i}, X_2, X_3) \qquad i = 1 \text{ to } n \tag{1}$$

$$L_i = f(I_i, F, X_{1i}, X_2, X_3) \qquad i = 1 \text{ to } n \tag{2}$$

$$I_i = f(F, X_{1i}, X_2, X_3, I_{i,t-1}) \qquad i = 1 \text{ to } n \tag{3}$$

$$F = f(X_2, X_3) \tag{4}$$

where X_{1i} denotes a set of parcel-specific characteristics, such as distance from house, parcel size, soil quality, and land rights; X_2 is a set of household-specific characteristics, such as family size, wealth, education, risk aversion, farm size, and fragmentation at inheritance; and X_3 is a set of village dummies to capture the effects of prices, the efficiency of local land, labor, credit, and food markets, and the available agroclimatic diversity. Subscript i indicates a particular parcel, and there are n parcels.

The yield of each parcel is determined in equation 1 by the use of inputs; the current stock of land improvements; the level of farm fragmen and parcel, household, and village dummies. Equations 2 and 3 determ level of direct inputs used on, and the current stock of land improvement to, the ith parcel, respectively. Both are determined by the level of farm fra tation, as well as by relevant parcel, household, and village variables, incl land improvements that have previously been made (I_{t-1}).

Unlike equations 1 to 3, farm fragmentation is determined in equation 4 a household rather than the parcel level. In keeping with our earlier discussion postulate that farm fragmentation depends on household (X_2) and village (variables. Because fragmentation is either exogenously determined by "supp side" factors or is a choice variable that predetermines the selection of speci parcels of land, the level of farm fragmentation enters the model as a predete mined variable (that is, in a recursive manner).

In this model fragmentation has direct and indirect effects on yields. It affects yields indirectly through the choice of input levels and land improvements. It also affects yields directly in equation 1 as might arise, for example, from lost output caused by theft or animal destruction, or from additional output as a result of superior matching of soils and crop types.

Because equation 4 does not include the current values of Y_i, L_i, or I_i, the model can be written in semi-reduced form. We are specifically interested in the semi-reduced form of the yield equation:

$$Y_i = f(X_{1i}, X_2, X_3, F) \qquad i = 1 \text{ to } n. \tag{5}$$

In this equation the farm fragmentation variable captures both the direct and indirect effects of fragmentation on yields.

Fragmentation is assumed to be predetermined before yields in the model. However, the estimated coefficient on the fragmentation variable in equation 5 will be biased if there are unobserved variables influencing both the level of fragmentation and current yields (for example, farmer skill). In order to overcome this difficulty, the appropriate estimation procedure is to first estimate equation 4 to obtain a fitted value for fragmentation ($\hat{F}$), and then to replace the actual values of farm fragmentation in equation 5 with the fitted values. This will give a consistent and unbiased estimate of both the direct and indirect effects of fragmentation on yield.

To estimate the fitted value for fragmentation, the following household-level model was used:

$$F_j = b_1 Z_j + b_2 V_k + e_j \tag{6}$$

where F_j is either the number of parcels or the Simpson index; Z_j is a vector of explanatory variables comprising the gender and formal education of the household head, whether the household head is migrant or native, the number of years the household head has farmed, inherited wealth, inherited farm size, and either the number of inherited parcels or the Simpson index measured on inherited

parcels; V_k is a binary variable indicating the location in the village k; b_1 and b_2 are vectors of coefficients to be estimated; and e_j is an error term. Because the Simpson index is bounded by zero and one, it is necessary to use truncation models to estimate equation 6. This problem does not arise when using the number of parcels as the measure of fragmentation because it is an unbounded variable.

Although these regressions will produce fitted values of fragmentation that are uncorrelated with the error term of equation 5, their usefulness depends on their correlation with the observed values of fragmentation. Fortunately, the correlations between the fitted and actual values of fragmentation were greater than 0.5 in all regions, and the explanatory power of the regressions was high. The following variables are consistently significant in all the regressions: inherited farm size (negatively related to fragmentation), the fragmentation of the inherited holding (positively related to fragmentation), and locational variables (mixed signs). The first two lend some support to supply-side arguments for fragmentation, whereas the latter could reflect supply- or demand-side factors.

The semi-reduced-form yield equation to be estimated was specialized to the following exponential functional form for parcel i of household j, located in village k:

$$\log Y_{ij} = \alpha V_k + \beta_1 \log X_{1ij} + \beta_2 \log X_{2j} + \beta_3 \hat{F}_j + \mu_j + e_{ij} \tag{7}$$

where, in addition to previously defined variables, μ_j is a household random effect satisfying $E(\mu_j) = 0$, $E(\mu_j^2) = \sigma_\mu^2$, and is independent across households, and e_{ij} is an error term. A random effects specification is used to allow for unobserved household effects when combining data for several parcels from each household.[1]

Multiple crops are often produced on a single parcel, and the corresponding areas devoted to each crop are not known. In order to aggregate yields of different crops, median prices are used so that the dependent variable is the value of crop output per hectare. The inclusion of multiple crops presents an additional problem. Some of the independent variables may have different effects on yields, depending on the cropping pattern. One can allow for varying effects in two ways. The first is to make separate regressions for each cropping pattern. Unfortunately, this procedure gives relatively small sample sizes (so that controlling for unobservable household effects is problematic) and, by giving crop-specific coefficients for all the independent variables, makes it difficult to

1. Because fitted values are used for F_j, these values will not be correlated with μ_j. However, it is possible that μ_j might be correlated with other parcel or household-specific variables, in which case the coefficients on some of those variables would be biased. This problem can be mitigated by using a fixed rather than a random effects specification. However, because F_j would then have to be dropped from the equation (it takes on the same value for all parcels operated by the jth household), we could not use this approach. This means that caution is required in interpreting coefficients on variables other than F that might be thought of as choice variables.

Table 4. *Selected Results from Parcel Yield Regressions for Three Regions in Ghana, 1987–88*

Variable	Estimate	t-statistic
Anloga		
Distance	0.056	2.02
Parcel size	−0.171	−5.48
Farm size	0.112	0.87
Simpson index	0.771	1.13
(Number of parcels)[a]	(0.043)	(1.18)
F-value	2123.0***	n.a.
R^2	0.91	n.a.
Wassa		
Distance—cocoa	0.103	1.49
Distance—cocoa and maize	0.243	2.31
Distance—cocoa and plantain	0.232	1.59
Distance—cocoa and kola	0.033	0.25
Parcel size—cocoa	−0.370	−5.43
Parcel size—cocoa and maize	−0.798	−4.41
Parcel size—cocoa and plantain	−1.257	−4.35
Parcel size—cocoa and kola	−0.161	−0.80
Farm size	0.100	0.52
Simpson index	1.038	0.74
(Number of parcels)[a]	(0.021)	(0.26)
F-value	365.0***	n.a.
R^2	0.81	n.a.
Ejura		
Distance—maize	−0.038	−0.40
Distance—yam	−0.407	−2.97
Distance—groundnuts	0.093	0.76
Parcel size—maize	−0.248	−3.08
Parcel size—yam	−0.341	−3.38
Parcel size—groundnuts	−0.371	−3.64
Farm size	0.202	1.31
Simpson index	0.271	0.55
(Number of parcels)[a]	(0.157)	(1.09)
F-value	627.0***	n.a.
R^2	0.83	n.a.

*** *F*-ratio that is significant at the 1 percent level.

n.a. Not applicable.

Note: Other variables included in the regression were characteristics related to parcel (cropping pattern, years since acquired, improvements made before acquisition); household (wealth per hectare, nonfarm income per capita, number of male equivalents per hectare); household head (gender, migrant status, occupation, farming experience, education); and location (village binary variables).

a. The reported coefficients were obtained using the Simpson index as the measure of fragmentation. Separate regressions were run using the number of plots as an alternative measure of fragmentation. Because the results did not change much, we have simply included the coefficients for the number of parcel variables in this row.

Source: Authors' calculations based on World Bank survey data.

summarize results for the fragmentation variables. If certain variables are not crop-specific, then a better method is to stack the separate cropping pattern yields into a single regression. Each crop-specific variable is then multiplied by its respective crop dummy variable. Distance to parcel, parcel size, and soil type are examples of crop-specific variables. Thus, our X_{1ij} vector in equation 7 is comprised of crop-specific and non–crop-specific variables. No household or farm variables were considered to be crop-specific.

V. The Results

The econometric model is used in this section to test whether fragmentation has a detrimental effect on land productivity. Specifically, our null hypothesis is that fragmentation is inefficient and reduces yields. Rejection of this hypothesis would not prove that fragmentation is demand-driven, because other factors may explain the lack of a negative relation between fragmentation and yield. But it would add credence to the demand-side explanation.

Ghana

In Anloga yield data are only available for shallot parcels, because other cropping patterns are rare. In Wassa there are enough observations for four cropping patterns: cocoa, cocoa and kola, plantain and cocoa, and maize and cocoa. In Ejura maize, yams, and groundnuts are included. Because each region produces distinct crops, there is no gain from combining them into a single regression; thus regressions are made separately for each region. Table 4 summarizes the key fragmentation results. The R^2s and the statistical significance of the regressions are good, yet neither the Simpson index nor the number of parcels (fitted values) has a significant effect on yields.[2] Thus we reject the null hypothesis that fragmentation is inefficient.

The distance to parcel was significantly related to yield in a few cases. In Anloga the distance to parcel had a positive impact on parcel yields. An explanation for this counterintuitive finding is that the more fertile of two primary growing areas, and that which is more prone to flooding, is located farther from the villages. The distance to parcel was also positively related to yields in Wassa, although not all the coefficients are significant. One explanation for the outcome is that the farther parcels represent the more recently developed lands. As such they have not lost as many nutrients from continuous cultivation as have other parcels. In Anloga this explanation does not hold, because all the available arable land was used many years ago.

In Ejura the only significant result was that the distance to parcels monocropped with yams negatively affected yields. Because distances to parcels are quite long in Ejura, it is likely that this finding reflects the inefficient use of labor

2. A similar lack of relationship was obtained when we used actual rather than fitted values of the Simpson index or the number of parcels.

Table 5. *Selected Results from Parcel Yield Regressions for Three Prefectures in Rwanda, 1987–88*

Variable	Pooled	Ruhengeri	Butare	Gitarama
Distance				
Beans	−0.02	−0.03	—	—
	(0.63)	(0.66)		
Coffee	0.05	—	0.10	−0.04
	(1.74)		(2.05)	(0.94)
Sorghum	0.05	—	0.08	0.06
	(1.54)		(1.37)	(1.16)
Sorghum and maize	−0.03	—	—	−0.06
	(0.58)			(0.89)
Sorghum and sweet potato	−0.02	—	0.03	—
	(0.36)		(0.41)	
Sorghum and other	−0.09	—	−0.16	−0.08
	(1.71)		(1.62)	(1.02)
Sweet potato	0.04	0.02	0.02	0.03
	(1.73)	(0.42)	(0.41)	(0.88)
Parcel size				
Beans	−0.40	−0.35	—	—
	(4.91)	(3.85)		
Coffee	−0.30	—	−0.21	−0.40
	(4.35)		(2.04)	(3.97)
Sorghum	−0.42	—	−0.52	−0.43
	(5.88)		(4.01)	(4.91)
Sorghum and maize	−0.50	—	—	−0.62
	(4.12)			(4.16)
Sorghum and sweet potato	−0.31	—	−0.33	—
	(2.85)		(2.48)	
Sorghum and other	−0.47	—	−0.53	−0.34
	(4.79)		(3.92)	(2.04)
Sweet potato	−0.42	−0.29	−0.32	−0.55
	(8.19)	(2.71)	(3.30)	(6.28)
Farm size	−0.03	−0.23	−0.10	0.07
	(0.33)	(0.76)	(0.52)	(0.21)
Simpson index	0.25	0.35	0.33	−0.03
	(0.64)	(0.32)	(0.44)	(0.02)
Number of parcels[a]	0.02	−0.01	0.03	−0.01
	(0.74)	(0.09)	(0.56)	(0.04)
F-value	21.01***	8.26***	8.38***	8.23***
R^2	0.56	0.53	0.60	0.68

— Not included in a particular regression.

*** F-ratio that is significant at the 1 percent level.

Note: Figures in parentheses are the absolute values of the *t*-statistics. Other variables included in the regression were characteristics related to parcel (cropping pattern, slope, topographical location, years since acquired, improvements made before acquisition); household (wealth per hectare, nonfarm income per capita, number of male equivalents per hectare); household head (gender, occupation, farming experience, education); and location (communal binary variables).

a. The reported coefficients were obtained using the Simpson index as the measure of fragmentation. Separate regressions were run using the number of plots as an alternative measure of fragmentation. Because the results did not change much, we have simply included the coefficients for the number of parcels variable in this row.

Source: Authors' calculations based on World Bank survey data.

and labor-intensive inputs on the more removed parcels. In all cases parcel size is negatively related to yield, again suggesting that problems in the labor market might be an important reason for fragmentation.

Rwanda

The analysis was confined to seven cropping patterns for which a sufficient number of parcel observations on output data were available: monocropped parcels of beans, sweet potatoes, sorghum, and coffee and sorghum multicropped with either maize, sweet potatoes, or other crops.

Table 5 summarizes the key fragmentation results for the pooled and individual prefecture-level regressions (many other variables were included in the regressions, and the full results are available from the authors). The regressions have relatively high R^2s and are statistically significant at the 1 percent confidence level. Nevertheless neither the Simpson index nor the number of parcels (fitted values) was significantly related to parcel yields in any of the regressions. Hence we again reject the hypothesis that fragmentation is inefficient.

While possible biases caution against undue reliance on some of the other coefficient estimates, farm size is not significantly related to yield in any of the regressions. The distance to parcel has a positive effect on yields for coffee in Butare and in the pooled sample. This probably indicates certain locational advantages—such as coffee parcels being located higher up the hillsides—rather than refuting the purely negative effect of distance on yields through inefficient use of labor and other inputs. Distance has a negative effect on yields of sorghum intercropped with "other" crops in the pooled regression, but the result is only weakly significant, and yields of other sorghum intercrops are not related to distance.

Parcel size is negatively associated with yield in all the regressions. To the extent that diseconomies in parcel size reflect problems in the labor market, these results add credence to the demand-side explanation for fragmentation.

VI. Fragmentation as a Risk Management Aid

Our econometric results refute the hypothesis that fragmentation is inefficient and suggest, but do not prove, that the private benefits from fragmentation are at least as large as the private costs. Although comprehensive analysis of these costs and benefits would require a much more structural and micro-oriented analysis than the available data permit, we can at least test to see if fragmentation confers benefits by reducing risk through less variable output.

A Model of Farm Fragmentation and Risk Spreading

The risk-reducing benefits of fragmentation can be approximated by the reduction in the variance of farm income obtained from cultivating several diverse parcels. A useful model has variations from mean yield arising from fluctuations at the village level, the farmer level, and the parcel level.

Assume that a single crop is produced at a fixed price, which, for convenien is normalized to unity. The yield for farmer h on parcel i in period t is given by

$$(8) \qquad Q_{hit}/A_i = (d_0 + \theta_t) + (g_h + \eta_{ht}) + (k_i + \gamma_{it})$$

where Q_{hit} is the output of the ith parcel, A_i is the area of the ith parcel, d_0 is the village mean yield for all parcels, θ_t is a random village-level effect, g_h is a fixed farmer effect, η_{ht} is a random (over time) farmer effect, k_i is a fixed parcel effect, and γ_{it} is a random (over time) parcel effect.

We assume that $E(\theta_t \eta_{ht^*}) = 0$ and $E(\theta_t \gamma_{it^*}) = 0$, for all i, h, t, t^* and that $E(\eta_{ht} \eta_{st^*}) = 0$ and $E(\gamma_{it} \gamma_{jt^*}) = 0$, for all $h \neq s, i \neq j, t, t^*$. Then, letting farmer h's total area be $A = \Sigma_i A_i$, total farm output is

$$(9) \qquad \Sigma_i Q_{hit} = (g_h + \eta_{ht} + d_0 + \theta_t) A + \Sigma_i A_i(k_i + \gamma_{it}).$$

The mean output is

$$(10) \qquad E(\Sigma_i Q_{hit}) = (g_h + d_0) A + \Sigma_i A_i k_i$$

and the variance is

$$(11) \qquad \mathrm{Var}(\Sigma_i Q_{hit}) = A^2 (\mathrm{var}(\eta_h) + \mathrm{var}(\theta)) + \Sigma_i A_i^2 \mathrm{var}(\gamma_i).$$

Or, assuming the γ_i are identically distributed with var(γ),

$$(12) \qquad \mathrm{Var}[(\Sigma_i Q_i)/A] = \mathrm{var}(\theta) + \mathrm{var}(\eta_h) + (\Sigma_i A_i^2/A^2)\,\mathrm{var}(\gamma).$$

The coefficient on var(γ) is equal to $1 - S$, where S is the Simpson index. Thus in this model risk, as measured by the variance of total farm income per hectare, declines linearly as the Simpson index increases.

Data and Estimation for Rwanda

Our risk analysis requires time series data on crop production. Such data are available for Rwanda in the National Agricultural Survey (NAS), collected by the Service des Enquetes et des Statistiques Agricoles of the Ministry of Agriculture. Data were obtained for a representative national sample of 1,015 farms. The data combine the results of a detailed household survey for 1986 with aggregate farm output information for six growing seasons (two a year from the second half of 1985 to the first half of 1988). The 1986 cross-section data include information on land fragmentation as measured by the Simpson index; location by prefecture, commune, and village; and soil fertility classification.

Output information is available for a variety of crops at the farm level, but there is no information on yields or on the allocation of crops between parcels. Estimates of the benefits of diversification of output for individual crops are not feasible without implausible assumptions, such as unchanging land use. For this reason output information for the various crops was aggregated using 1986 price weights to form an index of total output.

Unfortunately, information on the output of coffee, groundnuts, and soybeans is not available. This omission is not very serious for groundnuts and

sion from the measure of total output has the potential to introduce significant biases. Therefore attention was concentrated on Ruhengeri prefecture, with 105 observations, for which coffee output is known from the land rights survey to be a small proportion of total output (less than 2 percent).

Without parcel-level information, equation 12 can be used as the basis for a linear regression. As with equation 5, this is a semi-reduced-form equation. The dependent variable is the variance of income per hectare, and the independent variable is equal to one minus the Simpson index. The intercept is an estimate of var(θ). The slope coefficient is an estimate of var(γ). The error term is an estimate of var(η_h).

One problem in estimating equation 12 is that the fragmentation index may not have remained constant for all farms over the six growing seasons. Without time series data on fragmentation levels, little could be done to correct for any biases introduced by any changes. Fortunately, although short-term leasing and borrowing arrangements are quite common in Rwanda, the land rights study data do show that most parcels held on short-term arrangements are held by the same tenants for extended periods.

Preliminary analysis showed that variability of output was strongly associated with soil fertility in Ruhengeri. Somewhat surprisingly perhaps, output variability was higher for the more fertile soils. The Simpson index was also higher for farmers on the more fertile soils. As a result of this pattern, the raw correlation between the Simpson index and the degree of income variability is positive. If this correlation were (mis)interpreted as a coefficient estimate for equation 12, it would have the wrong sign.

The simplest way to correct for soil fertility is to include it as a separate regressor. The results are reported in the first row of table 6. The independent variable is the variance of income per hectare. The coefficient on fertility is positive and significant, and the coefficient on the Simpson index is negative but insignificant. It may be that because of possible simultaneity bias, the true effect of the Simpson index on variance is more significantly negative. A simultaneous system may exist because the level of fragmentation is observed during the

Table 6. *Regression Results for Risk Analysis for Ruhengeri Prefecture, Rwanda, 1985–88*

Adjusted sample	*Intercept*	*Simpson index*	*Soil fertility*	R^2
Complete data set	−5.94	−6.58	2.92	0.08
	(1.21)	(3.18)		
Good fertility level	1.27	−1.38	n.a.	0.10
	(2.55)			
Excellent fertility level	22.77	−21.27	n.a.	0.03
	(1.41)			

n.a. Not applicable.

Note: Figures in parentheses are the absolute value of the *t*-statistics.

Source: Authors' calculations based on World Bank survey data.

period in which the variance of yield is calculated. The bias is upward bec the farmer's choice of level of fragmentation would be a positive function of variance (for example, an increase in variance would cause farmers to raise t levels of fragmentation). The proportion of total variation explained by regression variables is small, but this is only to be expected in a model of kind.

The equation could be expanded to a more comprehensive treatment of fertility by including dummy intercept and slope shifters for the different types. It is simpler, however, to adopt the equivalent procedure of estimating regression in equation 12 separately for the different subsets of the data. Th different fertility levels were represented in the Ruhengeri prefecture. Howev there were only 12 farms with the lowest fertility level, so results are report only for the farms with good or excellent fertility levels. In both cases th Simpson index has the correct sign, but it is statistically significant only for th good soils. The estimated coefficient for the farms with excellent fertility is much higher, but the standard error is also very large.

VII. CONCLUSIONS

This article has shown that, although farm fragmentation is a pervasive feature of Ghanaian and Rwandan agriculture, it does not seem to have any adverse impact on the productivity of land. Many of the costs generally attributed to fragmentation (for example, travel time) are positively related to the distance between farmers' residences and their parcels. But most parcels are located within surprisingly short distances of the house, and, although fragmentation is positively (though rarely significantly) associated with increases in the average distance, the amount of travel time required is still only a matter of minutes. At the same time fragmentation increases the diversity of agroclimatic conditions available to the farmer, and this leads to more diversified cropping patterns. This can be beneficial for risk reduction, reducing peaks and troughs in labor demand and enhancing household food security.

The major implication of these findings is that, other things being equal, land consolidation policies are unlikely to increase land productivity significantly. Fragmentation is not as inefficient as widely assumed, and it offers farmers a tool for managing risk, seasonal labor shortages, and food insecurity when other alternatives might be more costly. A prime suspect for the lack of efficient alternatives is the inefficiency of many rural credit, labor, and food markets. While rigorous confirmation of these links would require more structural and micro analysis than the available data permit, policymakers could do worse than to focus their attention on improving the functioning of these markets and to avoid interventions in land markets (for example, restrictions on sales and rentals) that limit the ability of farmers to adjust optimally the extent of fragmentation (or consolidation) of their holdings over time.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Anthony, Kenneth R. M., Bruce F. Johnston, William O.Jones, and V. C. Uchendu. 1979. *Agricultural Change in Tropical Africa*. Ithaca: Cornell University Press.

Bentley, Jeffery. 1987. "Economic and Ecological Approaches to Land Fragmentation: In Defense of a Much-Aligned Phenomenon." *Annual Review of Anthropology* 16: 31–67.

Bauer, P. T., and Basil S. Yamey. 1957. *The Economics of Under-developed Countries*. University of Chicago Press.

Binns, Bernard O. 1950. "The Consolidation of Fragmented Agricultural Holdings." FAO Agricultural Studies 11. Food and Agriculture Organization, Rome.

Buck, G. L. 1964. *Land Utilization in China*. New York: Paragon.

Carter, Michael, and Peter Matlon. N.d. "An Empirical Characterization of Risk and Potential Household Differentiation in the Semi-Arid Tropics of West Africa." Department of Agricultural Economics, University of Wisconsin—Madison. Processed.

Charlesworth, Neil. 1983. "The Origins of Fragmentation of Landholdings in British India: A Comparative Examination." In Peter Robb, ed., *Rural India*. London: Carson Press Ltd.

Dahlman, Carl. 1980. *The Open Field System and Beyond*. Cambridge University Press.

Dorner, Peter. 1977. "Land Ownership and Tenure." In E. R. Duncan, ed., *Dimensions in World Food Problems*. Ames: Iowa State University Press.

Douglass, Larry W. 1969. *Death in Murelaga: Funerary Ritual in a Spanish Basque Village*. Seattle: University of Washington Press.

Downing, T. E. 1977. "Partible Inheritance and Land Fragmentation in an Oaxaca Village." *Human Organization* 36 (3, Fall): 235–43.

Farmer, B. H. 1960. "On Not Controlling Subdivision in Paddy Lands." *Transactions of the British Geographers* 28: 225–35.

Fenoaltea, Stefano. 1976. "Risk, Transaction Costs, and the Origin of Medieval Agriculture." *Explorations in Economic History* 13: 129–51.

Georgescu-Roegen, Nicholas. 1969. "The Institutional Aspects of Peasant Communities: An Analytical View." In Clifton R. Wharton, ed., *Subsistence Agriculture and Economic Development*. Chicago: Aldine Publishing Co.

Grigg, David. 1970. *The Harsh Lands: A Study in Agricultural Development*. New York: St. Martin's Press.

Holmberg, A. R., and H. F. Dobyns. 1969. "The Cornell Program in Vicos, Peru." In Clifton R. Wharton, Jr., ed., *Subsistence Agriculture and Economic Development*. Chicago: Aldine Publishing Co.

Hopper, W. David. 1965. "Allocation Efficiency in a Traditional Indian Agriculture." *Journal of Farm Economics* 47 (3, August): 611–24.

Hyodo, S. 1963. "Aspects of Land Consolidation in Japan." In Kenneth H. Parsons, Raymond J. Penn, and Philip M. Raup, eds., *Land Tenure*. Madison: University of Wisconsin Press.

Ilbery, Brian W. 1984. "Farm Fragmentation in the Vale of Evesham." *Area* 16 (2): 159–65.

Johnson, V. W., and Raleigh Barlowe. 1954. *Land Problems and Policies*. New York: McGraw-Hill.

King, Russell. 1977. *Land Reform: A World Survey*. London: G. Bell and Sons.

Leach, Robert. 1968. *Pul Elya, a Village in Celyon: A Study of Land Tenure and Kinship*. Cambridge University Press.

Leibenstein, Harvey. 1957. *Economic Backwardness and Economic Growth*. New York: Wiley & Sons.

Lipton, Michael. 1968. "The Theory of the Optimizing Peasant." *Journal of Development Studies* 4 (3, April): 327–51.

McCloskey, Donald N. 1975. "The Persistence of English Common Fields." In William Parker and Eric Jones, eds., *European Peasants and Their Markets*. Princeton University Press.

———. 1976. "English Open Fields as Behavior towards Risk." In Paul J. Uselding, ed., *Research in Economic History: An Annual Compilation*. Greenwich, Conn.: Jai Press.

McKinnon, R. I. 1973. *Money, Capital, and Economic Development*. Washington, D.C.: Brookings Institution.

McPherson, Malcolm F. 1982. "Land Fragmentation in Agriculture: A Selected Literature Review." Discussion Paper 141, November. Harvard Institute for International Development, Cambridge, Mass. Processed.

Migot-Adholla, Shem, Peter Hazell, Benoit Blarel, and Frank Place. 1991. "Indigenous Land Rights Systems in Sub-Saharan Africa: A Constraint on Productivity?" *The World Bank Economic Review* 5 (1): 155–75.

Netting, Robert. 1972. "Of Men and Meadows: Strategies of Alpine Land Use." *Anthropological Quarterly* 45: 145–57.

Quiggin, John. 1988. "Scattered Landholdings in Common Property Systems." *Journal of Economic Behavior and Organization* 9 (2): 187–202.

Sargent, F. O. 1952. "Fragmentation of French Farm Land: Its Nature, Extent, and Causes." *Land Economics* 28 (August): 218–29.

Thirsk, Joan. 1964. "The Common Fields." *Past and Present* 29: 3–25.

Thompson, Kenneth. 1963. *Farm Fragmentation in Greece*. Athens: Center for Economic Research.

Udo, Reuben K. 1965. "Disintegration of Nucleated Settlement in Eastern Nigeria." *Geographical Review* 55: 53–67.

Wolf, E. R. 1966. *Peasants*. Englewood Cliffs, N.J.: Prentice-Hall.

World Bank. 1978. *World Development Report 1978*. New York: Oxford University Press.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 2: 255-277

Project Evaluation and Uncertainty in Practice: A Statistical Analysis of Rate-of-Return Divergences of 1,015 World Bank Projects

Gerhard Pohl and Dubravko Mihaljek

This article analyzes the World Bank's experience with project evaluation for a sample of 1,015 projects by comparing estimated rates of return at appraisal with reestimated rates of return when construction works are completed, usually 5 to 10 years after appraisal. The analysis highlights the high degree of uncertainty in project analysis. A wide range of variables has been introduced to explain the observed divergence in appraisal and reestimated rates of return, but only a relatively small part of the divergence can be explained, even with the benefit of hindsight. Project analysis thus has to cope with a large degree of uncertainty, which the traditional methods of project evaluation and selection have not been able to reduce.

The World Bank's long history of project financing provides a unique opportunity to quantify the level of uncertainty in public sector investment projects in developing countries and to assess the effects of cost-benefit analysis on investment decisions. For projects whose costs and benefits are reasonably amenable to quantification, Bank staff calculate economic rates of return at appraisal and again at project completion (after construction works have been completed and the project begins normal operations). For more than 1,000 projects, economic rate of return estimates now are available for both appraisal and completion. The difference between these two estimates provides an interesting empirical measure of the uncertainty of development projects financed by the World Bank.

Cost-benefit analysis is a standard appraisal tool for selecting development projects at the World Bank and other development finance institutions. Several governments also have adopted these techniques in planning public investment projects. The World Bank broadly follows the Little-Mirrlees (1968, 1974) methodology, which expanded earlier approaches to cost-benefit analysis to take

Gerhard Pohl is with the Geneva Office of the World Bank. At the time of writing, Dubravko Mihaljek was with the Department of Economics at the University of Pittsburgh and the Economics Institute in Zagreb. Research for this study was initiated while Dubravko Mihaljek was a summer assistant in the Economic Advisory Staff of the World Bank. Helpful comments by the participants in a World Bank seminar and anonymous referees are gratefully acknowledged.

account of economic distortions typically prevailing in developing countri Squire and van der Tak (1975) refined the methodology further to take acco of income distribution effects, as did Reutlinger (1970) to take account of unc tainty. More recent generalizations and refinements of the theory of cost-bene analysis are reviewed, for example, in Drèze and Stern (1987) and Squi (1989).

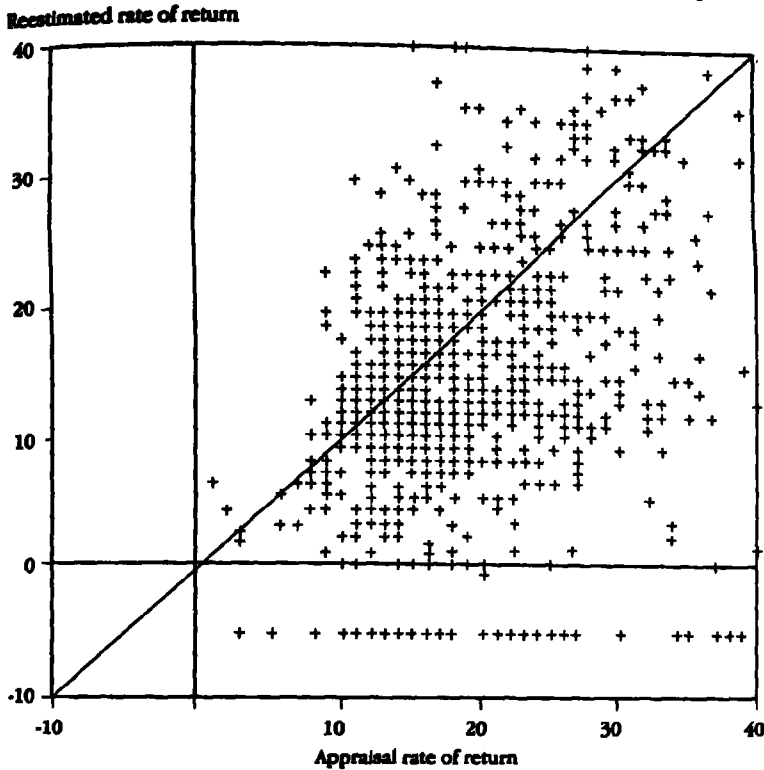
The theory of cost-benefit analysis provides a rigorous conceptual framework in which to evaluate public expenditure programs and investment projects (Little and Mirrlees 1991). Practical applications, however, depart quite substantially from these ideals because some of the key parameters are difficult to estimate in practice. In theory, for example, net discounted benefits at the accounting rate of interest would be the appropriate criterion to decide whether to carry out a particular project in the Little-Mirrlees framework. In practice it is quite difficult to estimate the accounting rate of interest reasonably accurately. Equating the accounting rate of interest with the (highest) rate that just exhausts available investment funds, as suggested early on by Little and Mirrlees, is conceptually clear and simple, but few, if any, developing countries have a comprehensive ranking of available public investment opportunities. Returns on past public sector investments may be misleading due to poor investment decisions or inappropriate economic policies. Project proposals by sectoral agencies, however, may be padded with optimistic assumptions and imply an exaggerated rate of return.

Since public sector and economywide rate-of-return estimates vary considerably across countries (depending on endowments and, more important, on policies), the Bank uses the internal rate of return calculated at shadow prices or, for short, the economic rate of return as an important, but not exclusive, decision criterion. With few exceptions the World Bank finances only projects that have an estimated economic rate of return of at least 10 percent at appraisal (in constant prices).

Of course projects financed by the World Bank are not necessarily representative of public sector investments in developing countries. Projects submitted by governments for Bank financing may primarily include projects with above-average rates of return and below-average risks. This may be particularly true in large countries (such as India), where Bank-financed projects represent only a very small part of the public investment program.

Although a considerable degree of uncertainty is to be expected in implementing development projects, the extent of revealed uncertainty in World Bank projects is striking. Estimates of economic rates of return at appraisal (AERR) are relatively poor predictors of reestimated rates of return at completion of construction works (RERR). Figure 1, in which each point represents rates of return for one project, shows this relation—most of the points in the figure are below the 45 degree line, that is, most projects have economic rates of return lower at completion than at appraisal.

The reestimated rates of return are not true ex-post rates of return because

Figure 1. ***Relation between Appraisal and Reestimated Rates of Return***

Note: Projects with appraisal or reestimated economic rates of return greater than 40 percent (about 8 percent of all projects) have been omitted for greater clarity.

Source: Authors' calculations.

they are made at the start-up of normal operations. In view of the long life of most investment projects, ex-post estimates can be made only after 10 or more years. However, when the reestimated rates are calculated, the effect of a number of risks already is known (such as investment costs, construction delays, and initial operating performance), and later costs and benefits are more heavily discounted. Because of the long life of most investment projects, the relation between ex-post rates of return and reestimated rates at completion of construction may be as loose as the relation between appraisal and reestimated rates of return.

This article analyzes the differences between appraisal and reestimated rates of return with statistical techniques and provides some initial interpretation of the results. The statistical analysis can capture only factors that are measurable and are applicable to all types of projects. This approach can provide only an

overview and cannot substitute for project performance audits at the project level. The rate of return of a copper project, for example, will depend strongly on whether actual copper price developments fulfill appraisal expectations. The closest our analysis gets to this is through inclusion of a composite real commodity price index on projects in broad sectors (for example, agriculture).

As discussed in the next two sections, the nature of our data set poses several challenges for statistical analysis. Two most serious ones are the heteroskedasticity of residuals and the censoring of reestimated rates of return. To avoid these problems, we have transformed the dependent and explanatory variables in such a way that the error term is constant, and we have used the maximum likelihood estimation technique within the Tobit regression framework. Tests performed confirm the validity of this approach and therefore give additional weight to our results.

I. The Data

This analysis is based largely on a data base maintained by the Bank's Operations Evaluation Department (OED), an independent unit reporting directly to the Executive Directors. The complete data set includes 2,200 projects for which project completion reports had been issued during 1974–87 by the respective project departments for audit by the OED on a sample basis. For slightly more than half of these projects, economic rates of return have been calculated by the staff at appraisal and project completion (start-up of normal operations). For the remainder of the projects in the data base (such as technical assistance and structural adjustment loans), economic rates of return were not available, primarily because quantification was deemed infeasible or unjustified. A few projects were excluded because other key variables were missing from the data set, or because other supplementary data were not available. The analysis thus was carried out with a final sample of 1,015 projects for which a complete data set was available. The sample selection is fairly objective and does not appear to bias the results. The only systematic omission is financial intermediation projects, which finance numerous small- and medium-scale industrial and agricultural projects. A similar analysis could, in principle, be carried out for financial intermediation projects, but would involve considerable effort to collect data.

The OED data base was augmented by supplementary variables that were believed to be important explanatory factors of project success, including the Bank's internal ranking of economic management performance by country as of 1978, a country-specific index of price distortions for the 1970s (Agarwala 1983), real commodity price movements, per capita income, and adult literacy rates. The price distortion index was available for 31 countries accounting for 612 projects. A separate analysis was undertaken for this smaller sample. Sectoral and regional dummy variables were introduced as proxies for project characteristics and management performance.

Table 1. *Summary Statistics for 1,015 World Bank Projects*

	Mean	Median	Maximum	Minimum	Standard deviation
Economic rate of return (percent)					
At appraisal	22	18	158	1	13
At project completion	16	14	128	−20	13
Total project cost (US dollar millions, current prices)					
At appraisal	86	34	3,193	1	185
At project completion	102	40	4,045	1	233
Nominal cost overrun (percent)	22	10	514	−89	46
Unexpected inflation (percent)	20	23	38	−2	7
Real cost overrun (percent)	−6	−11	394	−91	34
Time overrun (years)	2	2	16	−4	2
Time overrun (percent)	58	46	405	−68	56

Note: Project completion refers to start-up of normal operations.
Source: Authors' calculations.

Table 1 presents the main descriptive statistics for the data set and shows the wide variety of projects. Economic rates of return are derived, as explained above, as real internal rates of return at economic (shadow) prices. Project costs in the data base are in nominal U.S. dollars and range from about $1 million to more than $4 billion. Forty percent of the projects are in agriculture, 30 percent in transport, 20 percent in energy, and the rest in industry and urban development. The larger part of the Bank's industrial lending is intermediated through financial institutions. Unfortunately, average rates of return on financial intermediation lending are not systematically available.

Except for construction costs, the medians are fairly close to the averages for the data set. Constant-dollar cost data are not recorded in the data base, although they are available in the project files and have been used to calculate the economic rates of return. Rather than sifting through 1,015 voluminous project files, we have estimated implicit real costs from forecast and actual price developments as well as from other implementation data in the data base (see the appendix).

There is considerable variation in the appraisal rates of return, ranging from only 1 percent for a water supply project in Bombay to 158 percent for a seed project in India (which had a reestimated rate of return at project completion of only 11 percent). The wide range of rates of return both at appraisal and project completion (−20 to 128 percent) is perhaps surprising. In the more orderly world of economic theory and model-building, one usually assumes that rates of return converge within a fairly narrow range. Ninety percent of all projects have appraisal rates of return in the range of 10–40 percent, but only about half have reestimated rates of return within this range, which highlights the importance of uncertainty. The average rate of return of World Bank projects has behaved in a

more orderly way, averaging 22 percent at appraisal and 16 percent at project completion, with fairly small differences from year to year.

For about 5 percent of all projects the reestimated (internal) rates of return are negative. For many projects with negative rates of return these are given as −5 percent in the data base, presumably because negative internal rates are highly sensitive to small differences in assumptions and to truncation of the time horizon. The economic interpretation of negative internal rates of return in most cases is a zero rate of return (no increase in output for some large investment). Only if variable costs exceeded benefits (both at shadow prices) would one normally speak of "negative" rates of return, which typically would be due to domestic price distortions (negative value added at international prices).

Although the *average* reestimated rates of return are satisfactory (16 percent, versus 22 percent at appraisal), there are many projects with low returns (25 percent of all projects have reestimated rates of return below 10 percent, 14 percent have below 5 percent, and 8 percent of all projects have zero or negative rates of return). This suggests that considerable benefits could be obtained if the factors that lead to project failure could be identified. Defining project failure is not a simple matter. Some cutoff point for rates of return has to be adopted to distinguish successful from unsuccessful projects. In the Little-Mirrlees framework, that cutoff point would be relatively high (say 10 percent) due to the choice of the numeraire ("uncommitted public income," as opposed to consumption in the UNIDO methodology; see Dasgupta, Marglin, and Sen 1972 and Ray 1984).

World Bank projects have, on average, taken considerably more time to implement (six years) than expected at appraisal (four years), and project costs in U.S. dollars were, on average, 22 percent higher than estimated at appraisal, despite ample physical and price contingencies built into project cost estimates. Project cost overruns and implementation delays thus could be important factors in explaining project performance and the loose relationship between rate-of-return estimates at appraisal and project completion.

With the help of the derived real project costs, the reported nominal cost overruns were decomposed into two parts: unexpected changes in the general price level for capital goods and project-specific real cost increases. The latter could be due to an error in project cost estimates, unforeseen difficulties and expenditures, or increases in the scope of projects. Nominal cost overruns mostly are explained by unexpectedly high inflation during the period (primarily the 1970s), with actual prices being 20 percent higher than projected at appraisal. Perhaps surprisingly, the appraisal cost estimates were, on average, too high in *real terms*. Nominal cost overruns thus are primarily due to unexpectedly high inflation. Real cost variations range from −90 percent (probably largely due to cancellations of project components) to increases of nearly 400 percent (probably reflecting expansion in the scope of projects, rather than faulty cost calculations).

II. Statistical Methodology

The divergence between the appraisal and reestimated rates of return has been analyzed with two types of linear regression; the results from both types of regression are reported in section III. The first consists of regressing the reestimated rates of return at project completion on the appraisal rates of return and several other factors that are thought to influence project performance. Since both the appraisal and reestimated rates of return depend on the same set of other factors, this approach is best interpreted in terms of a "seemingly unrelated regression" model. (See Zellner [1962] for the original contribution; and Wallace, Duane, and Nawaz [1987] for an application similar to this article.)

One statistical problem with the standard ordinary least squares (OLS) estimation of this linear regression model is that residuals calculated from the above data set do not have uniform variance and zero correlation with one another. In the presence of heteroskedasticity the OLS estimator remains unbiased, but it no longer has minimum variance among all linear unbiased estimators. Also the usual formula for the variance-covariance matrix of OLS estimators is incorrect, and therefore the usual estimator of their variance is biased, implying that interval estimation and hypothesis testing using these estimators no longer can be trusted. Intuitively, one would expect larger time and cost overruns to be associated with larger discrepancies in the rates of return. For a wide range of projects the standard deviation of reestimated rates of return increases only moderately with appraisal rates of return (from 9 percentage points for projects with AERRs of 10–20 percent to 14 percentage points for projects with AERRs of 30–40 percent), but it jumps to 25 percentage points for a small number of projects with higher AERRs.

Several techniques exist to correct the standard errors of estimates (White 1980). We eliminate the problem by transforming the variables in such a way that the error term is constant. Thus in the second type of regression model the dependent variable is the percentage change in the reestimated rate over the appraisal rate of return [(RERR − AERR)/AERR]. This transformation eliminates heteroskedasticity, but at the cost of losing interesting information about the relation between the appraisal and reestimated rates of return. A third approach (not undertaken here) would be to eliminate projects with very high appraisal rates of return, because these may have an extraordinarily strong influence over the results. High AERRs often are due to major changes in expectations and usually involve comparatively small investments (for example, energy conservation, resource discoveries, and technological breakthroughs).

Besides heteroskedasticity, the nature of our data set gives rise to another statistical problem: censoring of reestimated rates of return. From table 1 it appears that the range of variation of reestimated rates of return is wide enough to make plausible the assumption of an approximately normal distribution of residuals. However, there is a considerable piling up (about 7 percent of proj-

ects) of RERRs at a cutoff point of −5 percent. This reflects the established practice in the World Bank whereby a project deemed to be a complete failure usually is assigned a −5 percent rate of return at completion.[1]

From an econometric point of view, the presence of the cutoff point implies that we are dealing with a censored sample, as some observations of the RERR that correspond to *known* values of time and cost overruns are not observable, being instead arbitrarily assigned the RERR of −5 percent. The difficulty with OLS estimation based on censored data samples is that the least squares estimators of regression parameters are biased and inconsistent, using either the entire sample or the subsample of complete observations. (For analysis of censored data samples see, for example, Maddala [1983] or Judge and others [1985].) These kind of data are best analyzed within the framework of the censored regression model, also known as the Tobit model (Tobin 1958). A number of techniques now exist to estimate Tobit models (see Amemiya 1984). To generate more efficient parameter estimates, we used the maximum likelihood technique, which yields estimators with several desirable asymptotic properties (see, for example, Greene 1990).

The Tobit regression model for our data sample is of the form:

$$(1) \qquad y_t = \begin{cases} x_t'\beta + e_t & \text{if } y_t > k, \\ 0 & \text{otherwise} \end{cases} \qquad i = 1, \ldots T - s$$

and the corresponding regression function is given by:

$$(2) \qquad E(y_t | x_t, y_t > k) = X_t'\beta + \sigma\epsilon_t$$

where y is a vector of dependent variables (RERRs or percentage changes thereof over the AERRs), X is a matrix of explanatory variables, β is a vector of unknown regression parameters, σ is an unknown scale parameter, e and ϵ are vectors of errors assumed to come from the standard normal distribution, k is the cutoff point (−5 percent for regressions where RERR is the dependent variable and −1 for regressions where the dependent variable is (RERR − AERR)/AERR)), and s observations out of T are unobservable.[2]

1. Another such practice is that projects with appraisal rates of return of less than 10 percent usually are not considered for approval. Theoretically, in the presence of this cutoff point the data set would be truncated: values of time and cost overruns and RERRs would be known only when AERRs at or above 10 percent were observed, so we could make no inference on the potential performance of projects that were not accepted for financing. However, in the data set there are 46 projects (4.5 percent of the total) that were approved even though they had AERRs of less than 10 percent, so information on normally "unobservable" projects actually is not missing.

2. The log likelihood function for this regression model is given by:

$$\ln L = -(n_1/2)[\ln(2\pi)+\ln\sigma^2] - (1/2\sigma^2)\Sigma_1(y_t-\beta' x_i)^2 + \Sigma_0\ln[1-\Phi(\beta' x_t/\sigma)]$$

where $\Phi(.)$ is the standard normal cumulative distribution function. The first two parts of this function correspond to the classical regression for noncensored observations; the last part corresponds to the relevant probabilities for the censored observations (see Amemiya 1973). MLE estimates of this regression model were computed using the LIFEREG procedure of the SAS statistical package, version 6.06.

III. The Models and Estimation Results

In the simplest model the reestimated economic rate of return is the dependent variable, and the appraisal economic rate of return is the explanatory variable. Additional models are presented with explanatory variables for cost overruns, implementation delays, and unexpected inflation. These explanatory variables are also used in alternative models with a transformed dependent variable. Two of the initial five models are estimated with additional explanatory variables to account for changes in primary commodity prices, economic management factors, and sectoral and geographic differences.

Five Initial Models

The simplest possible model relates only appraisal and reestimated rates of return and assumes that no other factors have been identified. Thus equation 3 represents model 1:

$$\text{RERR} = a + b\,\text{AERR} + \sigma u. \tag{3}$$

Models 2 and 3 expand on model 1 to include project cost overruns and implementation delays, which are intuitively linked with poor project performance. Model 2 introduces two variables from the data base: the nominal cost overrun and the time overrun, both in percent. Model 3 introduces two variables that decompose the nominal cost overrun into two components: unexpected inflation and real cost overruns.

Equation 4 gives the format for regressions with the transformed dependent variable (RERR − AERR)/AERR and the transformed explanatory variables $(X_t - X_{t-1})/X_{t-1}$:

$$(\text{RERR} - \text{AERR})/\text{AERR} = a + b' [(X_t - X_{t-1})/X_{t-1}] + \sigma u \tag{4}$$

where t denotes the relevant observation at the time of completion (of construction) and $t - 1$ the appraisal estimate of that same variable.

Maximum likelihood parameter estimates for models 1 through 5 are given in table 2, with summary statistics obtained from the OLS estimates of the regressions. Normal scale parameter σ does not have an intuitive economic interpretation, so its estimates are not reported. In all regressions where the dependent variable is RERR, the estimates of σ are on the order of about 12 percentage points and are statistically highly significant.

Model 1. The results for model 1 indicate that economic rates of return reestimated at project completion are, on average, considerably lower than appraisal estimates ($b = 0.44$). The intercept is quite large (5.88 percentage points), indicating that reestimated rates of return are somewhat higher, relative to appraisal estimates, for projects with low appraisal rates of return. A project with an appraisal rate of return of 10 percent has a reestimated rate of return approximately equal to its appraisal rate of return ($5.88 + 0.44 \times 10$), whereas

Table 2. Maximum Likelihood Estimates for the Initial Five Models

Explanatory variables and regression statistics	*Dependent variable: Reestimated economic rate of return*			*Change in reestimated economic rate of return relative to the appraisal economic rate of return*	
	Model 1	*Model 2*	*Model 3*	*Model 4*	*Model 5*
Explanatory variable					
Intercept	5.88	5.25	8.78	−101.42	−40.76
	(0.76)	(0.88)	(1.26)	(7.89)	(12.72
Appraisal rate of return	0.44	0.44	0.45	n.a.	n.a.
	(0.03)	(0.03)	(0.03)	n.a.	n.a.
Nominal cost overrun	n.a.	0.003	n.a.	0.07	n.a.
	n.a.	(0.009)[a]	n.a.	(0.11)[a]	n.a.
Time overrun	n.a.	0.009	0.012	0.01	0.03
	n.a.	(0.007)[a]	(0.007)	(0.01)[a]	(0.08
Unexpected inflation	n.a.	n.a.	0.23	n.a.	3.63
	n.a.	n.a.	(0.06)	n.a.	(0.70
Real cost overrun	n.a.	n.a.	0.01	n.a.	0.14
	n.a.	n.a.	(0.01)[a]	n.a.	(0.14
Regression statistics (OLS)					
Adjusted R^2	0.19	0.19	0.20	0.0012	0.03
F-statistic	240	80	64.6	0.59	11.02
Chi-square statistic	13.1[b]	26.8[b]	40.8[b]	9.38	17.4
(critical value for 1 percent level)	(9.2)	(21.7)	(29.1)	(15.1)	(21.7)
F-test of the regression	n.a.	0.7	5.0[c]	n.a.	28.8[d]

n.a. Not applicable.

Note: Sample size is 1,015 for all five models. Values in parentheses are the standard errors.

a. Not significant at the 5 percent level.

b. Presence of heteroskedasticity at the 1 percent level of significance.

c. Regression fit improved with respect to model 1.

d. Regression fit improved with respect to model 4.

Source: Authors' calculations.

a project with an appraisal rate of return of 30 percent has, on average, reestimated rate of return of 19 percent (5.88 + 0.44 × 30). As indicated by th low values of standard errors of estimates, both the intercept and the paramete estimate for AERR are statistically highly significant. However, the appraisal rat of return explains only 19 percent of the variance, which indicates a rather loos relation between rate of return estimates at appraisal and completion of con struction, already shown in figure 1.

Model 2. In model 2 parameter estimates for nominal cost overrun and tim overrun are small and statistically insignificant, thus indicating that nomina cost overruns and implementation delays do not seem to be major factors i explaining divergences in the rates of return.

Model 3. In model 3 the real cost overrun parameter remains low and statis tically insignificant, while the unexpected inflation variable is statistically signif icant. Since unexpected price increases have been expressed as negative number

(reduction in the real value of available project resources), the parameter estimate implies that for projects with (the average) unexpected increase in the price level of 20 percent, rates of return have been reduced by 4.6 percentage points. The results seem to suggest that real increases in project cost have had no systematic effect on rates of return of World Bank projects. Unexpected inflationary pressures have adversely affected the performance of Bank projects, perhaps because of relative price changes between capital-good inputs and project outputs.

Regression of model 3 also yields a statistically significant estimate of the time overrun variable, which, surprisingly, has the wrong sign. If, for example, it takes an average project 58 percent more time to be completed than forecasted at appraisal, one can expect that this would improve the RERR by about 0.7 percentage points. According to this result, the systematic bias toward underestimating the time needed for project completion may be based on the wrong intuitive assumption that long periods of implementation are bad for project performance. However, the modest positive effect that time overruns have on project performance must be weighted against the much bigger negative cost effects stemming from unexpected inflation.

Only the introduction of the decomposed cost overrun variables in model 3 improves the regression fit compared with model 1, as shown by the F-test for additional regressors. But the adjusted coefficient of determination ($\bar{R}^2$) improves by only one percentage point (from 19 to 20 percent). The chi-square statistic for the White test (see White 1980) indicates the presence of heteroskedasticity at the 1 percent test level.

Model 4. In model 4, which has the transformed dependent and explanatory variables, the estimated regressions have no heteroskedasticity at the 1 percent test level. However, the nominal cost and time overrun parameters are not statistically significant, and the estimated parameters have the wrong sign.

Model 5. In model 5 the decomposition of nominal cost overruns into unexpected inflation and real cost overruns helps somewhat, because the parameter estimate for the unexpected inflation variable now is statistically significant and has the expected (positive) sign (faster than expected inflation is a negative number). An unexpected price increase of 20 percent (the average for the sample) would thus give rise to a 73 percent discrepancy in the rates of return. However, although model 5 significantly improves the otherwise poor fit of model 4, the total explained variance of only 3 percent remains surprisingly low.

Project-specific real cost overruns thus do not seem to affect ex-post rates of return as adversely as one would expect. This may be due to the possibility that projects with large real cost overruns (up to almost 400 percent) reflect mostly expansions of projects, rather than errors in cost estimates. To the extent that project expansions lead to efficiency gains, one would expect improvements in the rate of return from such mislabeled "real cost overruns."

Table 3. *Maximum Likelihood Estimates for the Expanded Models*

	Dependent variable					
	Reestimated economic rate of return			*Change in reestimated economic rate of return relative to the appraisal economic rate of return*		
Explanatory variables and regression statistics	*Model 3A*	*Model 3B*	*Model 3C*	*Model 5A*	*Model 5B*	*Model 5C*
Explanatory variable						
Intercept	0.96	−11.47	0.2	−160.58	−194.05	−68.66
	(2.72)[a]	(5.77)	(0.001)[a]	(27.23)	(65.16)	(72.54)[a]
Appraisal rate of return	0.46	0.46	0.47	n.a.	n.a.	n.a.
	(0.03)	(0.04)	(0.04)	n.a.	n.a.	n.a.
Time overrun	0.015	0.01	0.01	0.07	−0.02	−0.002
	(0.007)	(0.009)[a]	(0.009)[b]	(0.08)[a]	(0.10)[a]	(0.10)[a]
Unexpected inflation	0.19	0.15	0.16	3.06	2.78	2.87
	(0.06)	(0.08)	(0.08)	(0.70)	(0.91)	(0.90)
Real cost overrun	0.006	0.003	0.004	0.09	0.18	0.17
	(0.01)[a]	(0.01)[a]	(0.01)[a]	(0.14)[a]	(0.17)[a]	(0.17)[a]
Unexpected change in commodity prices	8.19	9.49	9.6	118.95	103.58	105.46
	(2.35)	(3.04)	(3.0)	(27.44)	(34.72)	(34.24)
Economic management rating[c]	n.a.	0.42	n.a.	n.a.	6.07	n.a.
	n.a.	(0.36)[a]	n.a.	n.a.	(4.25)[b]	n.a.
Agarwala price distortion index[d]	n.a.	n.a.	−5.48	n.a.	n.a.	−63.46
	n.a.	n.a.	(1.40)	n.a.	n.a.	(16.54)
Log(GNP)	n.a.	1.84	2.06	n.a.	10.12	15.66
	n.a.	(0.88)	(0.80)	n.a.	(10.30)[a]	(9.32)
Adult literacy	n.a.	−0.06	−0.07	n.a.	−0.82	−0.93
	n.a.	(0.03)	(0.03)	n.a.	(0.34)	(0.33)
Regression statistics (OLS)						
Adjusted R^2	0.214	0.23	0.24	0.045	0.05	0.07
F-statistic	54.9	n.a.	n.a.	11.9	n.a.	n.a.
Chi-square statistic	63.7[e]	n.a.	n.a.	23.7	n.a.	n.a.
(Critical value for 1 percent level)	(37.6)	n.a.	n.a.	(29.1)	n.a.	n.a.
F-test of the regression	6.8[f]	4.5[h]	7.2[h]	14.0[g]	4.7[i]	7.9[i]
Sample size	1,015	612	612	1,015	612	612

n.a. Not applicable.
Note: Values in parentheses are the standard errors.
a. Not significant at the 5 percent level.
b. Not significant at the 5 percent level, but significant at the 10 percent level.
c. As of 1978, on a scale of 1 to 10; lowest actual rating is 2.
d. Ranges from 1.14 for Malawi (lowest distortion) to 2.86 for Ghana (highest distortion).
e. Presence of heteroskedasticity at the 1 percent test level.
f. Regression fit improved with respect to model 3.
g. Regression fit improved with respect to model 5.
h. Regression fit improved with respect to model 3A, adjusted for sample size.
i. Regression fit improved with respect to model 5A, adjusted for sample size.
Source: Authors' calculations.

Changes in Primary Commodity Prices

Since about 40 percent of the projects in the sample are agricultural projects, unexpected changes in commodity prices might explain a substantial part of the gap in the rates of return. We have chosen the ratio of the Bank's real commodity price index for 33 primary commodities (excluding energy) at project completion to the same index at the time of project appraisal as a measure of the extent of unexpected commodity price changes during project implementation. This measure is based on an adaptive expectations model of price expectations at the World Bank (see the appendix). Since the Bank's real commodity price index is deflated by the index of unit value of manufactured exports (MUV) of industrial countries, collinearity between the unexpected commodity price changes variable and the other price variables has been eliminated.

Table 3 gives the results of regressions of models 3 and 5 with this additional explanatory variable in the columns labeled model 3A and model 5A. The estimate for the unexpected commodity price changes variable is statistically significant, and its inclusion improves the fit of the regression, as measured by the F-test, with respect to models 3 and 5. The explained variance ($\bar{R}^2$) increases by about 1.5 percentage points, which is quite respectable compared to regressions with other variables, but the unexplained variance nevertheless remains very large.

The unexpected commodity price changes are measured as an index number, so no change corresponds to the index value of 1.0, and a 10 percent change corresponds to the index value of 1.1 or 0.9. The parameter estimates imply that an unexpected decline in commodity prices by 10 percent would reduce the rate of return by 0.8 percentage points [8.19 × (1.0–0.9)] in model 3A, or 12 percent (118.95 × 0.1) in model 5A. In model 5A the high estimated values of the intercept term (−161) and the commodity price parameter (119) actually must be set against each other for the zero expected price change to give, approximately, the intercept term from model 5 (−41). The same would hold true of model 3A if the intercept estimate was statistically significant.

A similar analysis was carried out for agricultural projects, using a real agricultural commodity price index, and the results were analogous. The use of individual commodity price indexes (such as a coffee price index for coffee projects) probably would show the greater sensitivity of some types of projects to specific commodity price changes, but the number of observations is too small to permit much further disaggregation. Also many agricultural projects are multi-purpose projects (for example, irrigation) for which the broad commodity price index may be more useful.

Economic Management Factors

A second set of factors that could help explain some of the divergence in rate-of-return estimates between the appraisal and completion of construction are the country-specific factors, such as the human resource endowment, the type of

economic policies pursued by the government, and the efficiency of public administration. Of course these are complex factors that are not easily measurable. We have to make do with a few quantitative indicators, such as adult literacy, per capita income, an index of price distortions for the 1970s (Agarwala 1983), and the Bank's internal ranking of the quality of government economic policies and management (as of 1978, taken as representative for the 1970s).

These factors should have been taken into account by project evaluators and factored into the appraisal estimate of the rate of return and, more important, into project design (for example, the extent of project management services provided by expatriates, which is one way of ensuring the success of the project). The parameter estimates for these variables thus need to be interpreted as the degree to which project evaluators did *not* sufficiently take account of these factors. In all cases it can be reasonably assumed that project evaluators were aware of these country-specific factors at the time of appraisal. Only in the case of the Agarwala price-distortion index could one possibly argue that there is some benefit of hindsight at work, because the extent of price distortions and their negative consequences may not have been fully appreciated. But Agarwala's index is based mostly on relatively easily available economic data that (at least in their raw form) were already available at appraisal. Moreover the Bank's internal rating of economic management performance is fairly closely related to the price distortion index (the coefficient of correlation between the two ratings is −0.67).

The results of regressions with country-specific variables added (models 3B, 3C, 5B, and 5C) are shown in table 3. The implementation delay and decomposed cost-overrun variables from models 3A and 5A have been retained. The Bank's economic management rating is an explanatory variable in models 3B and 5B; the Agarwala price distortion index is in models 3C and 5C. Most of the new variables are statistically significant.

The Agarwala price-distortion index performs statistically considerably better than the Bank's rating of economic management performance. This is surprising, since the latter is based on the same economic data *plus* management's judgments based on qualitative insights. Apparently the relatively simplistic procedure of adding up price distortions works better than a careful review process using qualitative judgments. In all model specifications, replacing the economic performance rating with the price distortion index results in statistically significant parameter estimates, a higher $\bar{R}^2$, and considerably higher values of F-tests for inclusion of the new regressors. For the actual range of the price distortion index (from 1.14 to 2.86), the parameter estimates imply a 9.4 [(2.86−1.14) × (−5.48)] percentage points (model 3C) lower RERR in a country with high price distortions (such as Ghana during the 1970s), compared with a country with low price distortions (such as Malawi). The adverse effects on project performance of government interventions through price controls, high tariffs, import restrictions, and so forth thus have been considerably underestimated in World Bank project appraisals.

Poor economic management and price distortions explain only about 2 percent of the rate-of-return gap, however, inching the total explained variance (model 3C) to only 24 percent. The level of income and adult literacy variables have been introduced as (albeit crude) indicators of the human capital stock, and their parameter estimates are statistically significant. Thus if per capita income decreases from $1,000 to $500, ex-post rates of return are lower by about 1.4 percentage points (model 3C), or about 11 percent (model 5C).[3] This suggests that the Bank's project evaluators have tended to overestimate project implementation capabilities in the poorest countries. Surprisingly, the parameter estimate for the adult literacy variable has the wrong sign, indicating that reestimated rates of return are lower in countries with higher adult literacy rates (for similar projects and levels of income). This can be explained by the fact that countries with high rates of literacy tend to engage in projects involving more sophisticated technology, which brings higher rates of return, but at higher risk, so that the rate-of-return discrepancy also is greater.

Sectoral and Geographic Differences

There are several ways to analyze the differences between the various types of projects. One is to introduce dummy (0,1) variables comparing different groups of projects. Another approach would be to run the same set of regressions on different sectoral or geographic subsets of projects to see whether there are statistically significant differences in parameter estimates. Table 4 presents estimates of regressions with both sectoral and regional dummy variables added to models 3B and 5B, now labeled 3D and 5D, respectively. Although it would have been preferable to use models 3C and 5C instead of 3B and 5B (because the Agarwala index performs better), this would have limited the sample to only 31 countries and 612 projects, instead of the entire sample of 1,015 projects. Agriculture and South Asia were selected as the standard to which other sectors and regions are compared, so parameter estimates for sectoral and regional dummy variables indicate how, for example, the energy projects or projects in the Mediterranean region perform relative to agricultural projects in South Asia.

The explanatory power of both regressions increases considerably after dummy variables are introduced (from an adjusted $\bar{R}^2$ of 0.21 to 0.31 for model 3D and from 0.03 to 0.12 for model 5D). These variables thus contribute more to the improved regression fit than all the previously introduced variables together (except for the appraisal rate of return in models 1 to 3). There thus appear to be clusters of projects with similar characteristics and problems.

Parameter estimates for sectoral dummy variables show that projects in our data sample fall roughly into two categories. Since the estimates for the intercept (that is, agriculture), energy, and industry all are insignificant, the results of model 3D indicate that, other things being equal, projects with an appraisal rate

3. Calculated as 2.06 (ln 1,000 − ln 500) = 1.43; and 15.66 (ln 1,000 − ln 500) = 10.85, respectively.

Table 4. *Regressions with Sectoral and Regional Dummy Variables*

Explanatory variables and regression statistics	*Dependent variable* — *Reestimated economic rate of return* — Model 3D	*Change in reestimated economic rate of return relative to the appraisal economic rate of return* — Model 5D
Explanatory variable		
Intercept	0.45[a]	−175.15
Appraisal rate of return	0.43	n.a.
Time overrun	−0.007[a]	−0.07[a]
Unexpected inflation	0.12	2.55
Real cost overrun	0.006[a]	0.12[a]
Unexpected change in commodity prices	6.45	90.49
Economic management rating	0.83[a]	6.53
Log(GNP)	−0.50[a]	11.94[a]
Adult literacy	−0.04[b]	−0.73
Sectoral dummy variable		
Energy	0.96[a]	27.41
Transport	7.62	63.58
Industry	−0.88[a]	−29.51[b]
Urban	9.51	−8.55[a]
Regional dummy variable		
East Africa	−12.54	−94.82
CFA countries	−8.02	−52.88
Other West Africa	−9.92	−96.87
East Asia	−3.64	−24.24[a]
Mediterranean	−6.92	−62.33
Latin America	−7.11	−63.04
Regression statistics (OLS)		
Adjusted R^2	0.309	0.121
F-test of the regression.	13.6[c]	8.1[d]

n.a. Not applicable.

Note: Sample size 1,015 projects for both models.

a. Not significant at the 5 percent level.

b. Not significant at the 5 percent level, but significant at the 10 percent level.

c. Regression fit improved with respect to model 3B.

d. Regression fit improved with respect to model 5B.

Source: Authors' calculations.

of return of, say, 20 percent that are undertaken in these three sectors are expected to have a RERR of about 9 percent (20 × 0.43), or 11 percentage points below the estimate. However, transport projects are expected to have a RERR about 4 percentage points below the estimate (7.62 + [20 × 0.43]), and urban development projects about 2 percentage points below the estimate (9.51 + [20 × 0.43]). This pattern is roughly confirmed by the results of model 5D, where the greatest rate-of-return discrepancy is for industry, there is less discrepancy for agriculture and energy projects, and transport projects again have the lowest discrepancy in economic rates of return. (Correct interpretation of the figures in

table 4 is more complicated, though, because we neglected the term $\sigma\epsilon_i$ [(equation 2)]; the scale parameter for model 3D is 11.4, and for model 5D it is 113.)

Reestimated rates of return for projects in the transport and urban development sectors thus are generally closer to the appraisal rates of return than are projects in agriculture, energy, and industry. This pattern probably reflects two factors: the relatively simple technology and organization of transport and urban development projects compared with industrial and energy projects and the effect of international markets on industrial and agricultural projects. Projects producing traded goods seem to be exposed to a higher degree of downside risks, and this may be related to international competition (that is, competitors in other countries may be more productive, and this may lead to lower prices for outputs and sharply lower returns).

Parameter estimates for regional dummy variables all have a negative sign, implying that reestimated rates of return are highest in South Asia (the standard of comparison), followed by the projects in East Asia, with slightly lower (3.6 percent) rates of return. Projects in Latin America, the Mediterranean, and Francophone Africa (CFA) are next on the list, while projects in East and West Africa (other than the CFA zone) have performed particularly poorly. The better performance of CFA members compared with other African countries points to the importance of the institutional framework and, in particular, the conservative monetary policies. The project implementation performance in CFA member countries during the 1960s and 1970s is comparable with the average of developing countries in the Mediterranean and Latin America.

An analysis of failed projects indicates that out of 80 total project failures (that is, negative rates of return at project completion), 27 are in East Africa, with Tanzania alone accounting for 11. Failed projects are concentrated largely in agriculture, as nearly two-thirds of all project failures worldwide have been agricultural projects, particularly complex new-style area or rural development projects started in the mid-1970s (table 5). Agricultural projects in Sub-Saharan Africa have an unacceptable failure rate, with one-half of all projects in East Africa and more than one quarter of all projects in West Africa yielding reestimated rates of return below 5 percent. There is a strong distinction in West Africa between CFA members and other countries.

A more informative approach to the analysis of the regional performance of projects is to run regressions for each sector separately (table 6). Compared with the combined sample (table 4), the disaggregated regressions by sector have fewer statistically significant parameters, and parameter estimates for some variables are very different from sector to sector. For the appraisal rate of return, the parameter estimates are relatively high for infrastructure projects (0.61–0.67) and low for agricultural and industrial projects (0.21–0.25). Thus downside risks are larger (or have been underestimated) in the directly productive sectors. Unexpected movements in primary commodities (excluding energy) seem to have affected industrial projects even more than agricultural projects (most of the industrial projects in the sample are import-substituting raw materials projects, such as in the fertilizer industry).

Table 5. *Project Failures by Region and Sector*

	Percentage of projects with reestimated rates of return below		
	10 percent	*5 percent*	*0 percent*
All projects	*25.2*	*13.6*	*7.9*
East Africa	41.1	27.9	17.1
West Africa			
CFA member countries	21.8	18.2	10.0
Other West Africa	37.5	19.6	16.1
Mediterranean	29.1	14.3	7.4
Latin America	25.1	10.2	5.1
South Asia	14.7	6.4	1.1
East Asia	12.5	5.2	4.2
Agricultural projects	29.0	19.8	12.7
East Africa	61.4	52.9	37.1
CFA members	34.0	26.0	16.0
Other West Africa	45.8	33.3	29.0

Source: Authors' calculations.

Table 6. *Regressions by Sector*

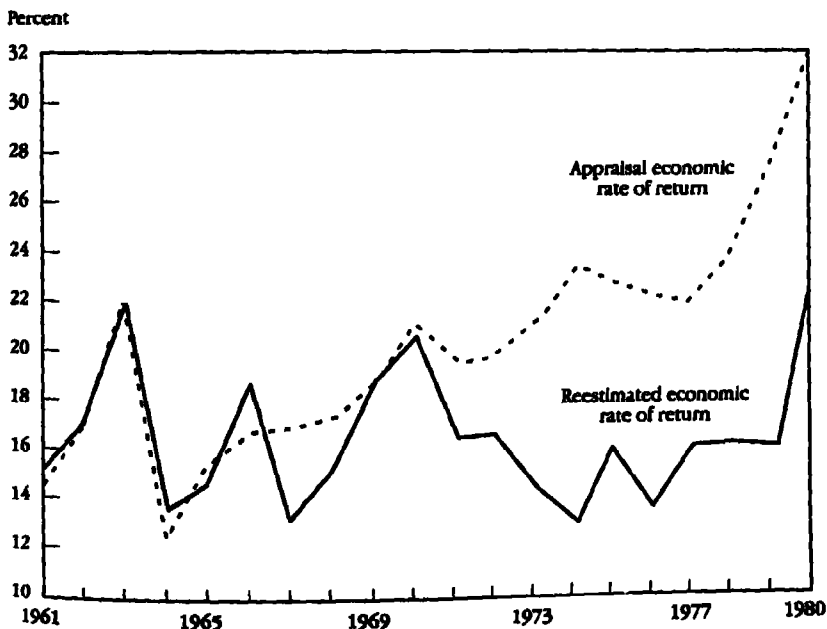
	Dependent variable: RERR *(model 3D)*			
	Agriculture	*Energy*	*Transport*	*Industry*
Explanatory variable				
Intercept	6.66[a]	10.26[b]	7.36[a]	−15.55[a]
Appraisal rate of return	0.21	0.61	0.67	0.25
Time overrun	−0.02[a]	0.002[a]	−0.01[a]	−0.02[a]
Unexpected inflation	0.11[a]	0.29	0.11[a]	−0.05[a]
Real cost overrun	0.03	−0.02[a]	−0.01[a]	−0.04
Unexpected change in commodity prices	11.12	1.92[b]	4.36[a]	15.72
Economic management rating	0.61[a]	1.17	0.52[a]	−1.72
Log (GNP)	0.82[a]	−1.21[a]	0.62[a]	2.35[a]
Adult literacy	−0.09	0.03[a]	−0.03[a]	0.10[b]
Regional dummy variable				
East Africa	−18.18	−5.09	−12.83	−14.06
CFA countries	−12.64	2.91[a]	−10.38	n.a.
Other West Africa	−17.37	−7.89	−6.33[b]	n.a.
East Asia	−3.29[b]	−4.42[b]	−7.65	5.24[b]
Mediterranean	−6.34	−7.46	−7.38[b]	−4.11[a]
Latin America	−8.34	−6.76	−6.97[b]	−2.81[a]
Regression statistics (OLS)				
Adjusted R^2	0.25	0.30	0.32	0.33
Adjusted R^2 without regional dummies	0.11	0.29	0.29	0.03
Sample size	411	216	310	56

a. Not significant at the 5 percent level.
b. Not significant at the 5 percent level, but significant at the 10 percent level.
Source: Authors' calculations.

Unexpected inflation seems to have affected particularly the more capital intensive infrastructure projects but may also reflect delayed adjustments in government price regulations in the electric power sector, where the Bank's methodology does not conform to the principles of cost-benefit analysis. The economic management rating also shows some perplexing sectoral differences, with a negative parameter estimate for industrial projects, which may be the consequence of a few conspicuous project failures in countries with high performance ratings. A separate analysis for agricultural projects showed that the Agarwala price distortion index performs dramatically better than the economic performance ranking in that sector.

IV. Trends over Time

An analysis of appraisal and reestimated rates of return by year of approval shows that the gap in the rate of return has increased considerably over time (figure 2). Projects appraised in the 1960s showed little difference between average rates of return at appraisal and completion, and annual variations were tracked quite closely. Appraisal and reestimated rates of return started to diverge in the early 1970s. The main cause appears to have been the increasing optimism

Figure 2. *Rates of Return by Year of Approval, 1961-80*

Source: Authors' calculations.

of project evaluators, with average appraisal rates of return rising from about 16 percent for projects evaluated in the mid-1960s to 20–25 percent for projects evaluated in the mid- and late 1970s. By contrast, average rates of return at project completion showed a persistent downtrend for projects appraised during 1970–76 before recovering again for projects appraised around 1980 (and evaluated in 1985–87).

The downtrend in average reestimated rates of return for projects appraised in the early 1970s was most likely due to external circumstances, that is, the recession and low commodity prices at the time these projects were completed in the late 1970s and early 1980s. The increasing rates of return at appraisal during the 1970s probably reflect internal Bank factors, including the shift in Bank lending from infrastructure to agriculture and industry. In terms of average outcomes at project completion, the vastly expanded lending program of the 1970s does not compare too unfavorably. Reestimated rates of return for the 1970s are not very different from those for the 1960s. The sharp increase in reestimated returns for projects appraised in 1980 must be interpreted with caution, because it includes only a small percentage of projects of that appraisal year.

V. Conclusions

The statistical analysis of rates of return estimates before and after completion of project construction provides several interesting insights. First, it points to the large degree of uncertainty surrounding the rate-of-return estimates. Second, World Bank appraisal estimates of rates of return are biased, that is, too optimistic. If this degree of optimism is shared by other project evaluators, one should expect that the discount rate that just rations investment projects to the funds available exceeds the ex-post rate of return by a considerable margin. The analytical treatment of project risks thus deserves more attention in practice. Anderson and Quiggin (1990), for example, argue that project implementation variables usually enter project analysis on a "no surprises" basis, corresponding to the modal value of the distribution of possible outcomes. Since surprises are mostly unpleasant, the probability distribution of project implementation outcomes is skewed (a longer tail in the downside direction). If one were to allow for the skewed distribution ("bad surprises"), one could correct the bias in the estimate.

However, factors that have conventionally been associated with this bias (such as cost overruns and implementation delays) seem to explain only a very small part of the unexpected changes in project performance (measured by the rate of return gap). Interestingly, uncertainties seem to be higher in the directly productive sectors (such as agriculture and industry), where rates of return can be altered through external market forces or domestic policy shocks. Rate-of-return estimates seem to be more stable for infrastructure projects.

As an alternative to correcting modal estimates of implementation variables

for "bad surprises," one could set different minimum rate-of-return criteria for different types of project (for example, 10 percent for transport, but 15 percent for agricultural and industrial projects), based on observed divergences in rates of return.

The analysis also has pointed to the importance of the policy environment for successful project implementation. The economic management rating and price distortion variables both indicate that project evaluators did not take the adverse effects of poor economic policies at the macroeconomic level sufficiently into account. More puzzling though, is the fact that regional dummy variables also seem to operate partly as economic management variables and have considerably more explanatory power than direct indicators of the quality of economic management and institutions.

The fact that projects in CFA member countries seem to perform almost as well as those in other regions shows that the high failure rate of projects elsewhere in Sub-Saharan Africa seems to be related primarily to policies and institutions. However, the better performance of projects in CFA countries during the 1960s and the 1970s is no guarantee that this will be repeated during the 1980s, because the external competitiveness of the CFA zone has considerably deteriorated.

The analysis of observed rate-of-return divergences raises more questions than it can answer. The high degree of revealed uncertainty also raises the question whether, and what kind of, improvements in the methodology will contribute to better investment decisions.

Appendix. Derivation of Project Data in Constant Prices

Project cost estimates for World Bank projects are made in current U.S. dollars, since this is the unit of account for the Bank. Appraisal estimates for a project are made on the basis of prevailing prices at the time of appraisal, a forecast of price changes for internationally traded capital goods (in terms of U.S. dollars), and the projected expenditure (disbursement) profile. Project cost estimates also include a physical contingency for unexpected expenditures. Project costs at project completion are, similarly, the sum of annual expenditures in (actual) current prices ("mixed year dollars").

In periods of unexpectedly high inflation (or for projects with significant implementation delays) reestimated nominal project costs sometimes are substantially higher than appraisal estimates, but real project costs may not have increased at all. To separate nominal from real cost overruns, we derived real (constant price) cost estimates for each project for both appraisal and project completion, by deflating yearly project expenditures with the projected and actual price index for capital goods (the Bank's manufactured unit value (MUV) index for exports of manufactured goods of industrial countries).

Although forecasts for the MUV index were available for the past 10 years, we did not have earlier forecasts and had to estimate the price contingency vectors.

Visual inspection of price forecasts for the past 10 years suggested that the Bank's price forecasts followed a pattern of "adaptive expectations": projections seemed to be based on recent price trends. Several adaptive expectations models were tested, and we found that the projections were best approximated by a five-year moving average adaptive expectations model.

The first stage prediction of the MUV index was made by calculating the five-year moving average:

$$\text{MUV}' = \frac{[\text{MUV}_{t-1} + \ldots + \text{MUV}_{t-5}]}{5}. \tag{A-1}$$

The moving average values (MUV′) were then used to calculate the average deviation from the actual value of the index (MUV′ − MUV)/MUV, and this estimate was used as a correction parameter, β, in the adaptive expectations model:

$$\text{MUV}^*_t = \text{MUV}'_{t-1} + \beta(\text{MUV}_{t-1} - \text{MUV}'_{t-1}), \qquad 0 < \beta < 1. \tag{A-2}$$

That is, the forecasting error for the previous period, $\text{MUV}_{t-1} - \text{MUV}'_{t-1}$, is corrected with a fraction β (the average error), thereby improving upon ("adapting to") the first-stage forecast.

The real cost at appraisal price projections, p^a, is then:

$$X^a(p^a) = \sum_{j=1}^{t^a} \frac{c^a/t^a}{p^a_j} \tag{A-3}$$

where X^a is real cost at appraisal, t^a is projected duration of project implementation, c^a is nominal cost estimate at appraisal, and p^a_j is projected price vector at appraisal. The real cost at actual prices is

$$X^c(p^c) = \sum_{j=1}^{t^c} \frac{c^c/t^c}{p^c_j} \tag{A-4}$$

where X^c is real cost based on actual nominal expenditures (c^c), actual implementation duration (t^c), and actual prices (p^c).

The real cost overrun is then $[X^c(p^c) - X^a(p^c)]/X^a(p^c)$, and the unexpected inflation, as defined in the article is $[X^a(p^c) - X^a(p^a)]/X^a(p^a)$. Unexpectedly high inflation ($p^c > p^a$) is a negative number according to this definition. This is reflected in the minus sign before the percentage sign in the tables with the statistical results.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Agarwala, Ramgopal. 1983. *Price Distortions and Growth in Developing Countries.* World Bank Staff Working Paper 575. Washington, D.C.

Amemiya, Takeshi. 1973. "Regression Analysis When the Dependent Variable Is Truncated Normal." *Econometrica* 41: 997–1016.

———. 1984. "Tobit Models: A Survey." *Journal of Econometrics* 24: 215–22.

Anderson, Jack, and John Quiggin. 1990. "Uncertainty in Project Appraisal." Paper presented at the World Bank Conference on Development Economics, Washington, D.C. Processed.

Dasgupta, Partha, Stephen Marglin, and Amartya Sen. 1972. *Guidelines for Project Evaluation.* New York: UNIDO.

Drèze, Jean, and Nicholas Stern. 1987. "The Theory of Cost-Benefit Analysis." In Alan Auerbach and Martin Feldstein, eds., *Handbook of Public Economics.* Vol. II. Amsterdam: North Holland.

Greene, William H. 1990. *Econometric Analysis.* New York: Macmillan.

Judge, George, R. Carter Hill, William Griffiths, Tsoung-Chao Lee, and H. Lütkepohl. 1985. *The Theory and Practice of Econometrics.* Second Edition. New York: Wiley.

Little, Ian, and James Mirrlees. 1968. *Manual of Industrial Project Analysis.* Vol. 2. Paris: OECD Development Centre.

———. 1974. *Project Appraisal and Planning for Developing Countries.* New York: Heinemann.

———. 1991. "Project Appraisal and Planning Twenty Years On." *Proceedings of the World Bank Annual Conference on Development Economics, 1990.* Washington, D.C.

Maddala, G. S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press.

Ray, Anandarup. 1984. *Cost-Benefit Analysis: Issues and Methodologies.* Baltimore, Md.: Johns Hopkins University Press.

Reutlinger, Shlomo. 1970. *Techniques for Project Appraisal under Uncertainty.* Baltimore, Md.: Johns Hopkins University Press.

Squire, Lyn. 1989. "Project Evaluation in Theory and Practice." In Hollis Chenery and Trent Srinivasan, eds., *Handbook of Development Economics.* Vol. II. Amsterdam: North Holland.

Squire, Lyn, and Herman G. van der Tak. 1975. *Economic Analysis of Projects.* Baltimore, Md.: Johns Hopkins University Press.

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables." *Econometrica* 26: 24–36.

Wallace, T. D., Paul Duane, and Tawhid Nawaz. 1987. "Agricultural Project Economic Rates of Return: A Statistical Analysis." Operations Evaluation Department, World Bank, Washington, D.C. Processed.

White, Halbert. 1980. "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48: 817–38.

Zellner, Arnold. 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregations Bias." *Journal of the American Statistical Association* 57: 348–68.

THE WORLD BANK ECONOMIC REVIEW, VOL 6, NO. 2: 279-305

The Business Cycle Associated with Exchange Rate-Based Stabilizations

Miguel A. Kiguel and Nissan Liviatan

This article examines the effects of disinflation on economic activity in countries characterized by chronic inflation. Such countries have a long history of inflation at rates exceeding those in industrial countries as well as labor and capital markets that have adjusted to function in an inflationary environment. A sample of disinflation programs in several Latin American countries and in Israel demonstrates that stabilization efforts in countries with chronic inflation often do not induce the usual Phillips curve tradeoff in the medium run. Specifically, stabilization programs that use the exchange rate as the main nominal anchor are often associated with a business cycle that begins with a boom and ends with a recession. Stabilization programs that use money supply as the nominal anchor generally induce the expected Phillips curve result: lower inflation is accompanied by a recession after the program is implemented.

This article examines the effects of disinflation on economic activity in countries characterized by chronic inflation, a term coined by Pazos (1972). In these countries there is a long history of inflation at rates exceeding those in industrial countries, and labor and capital markets have adjusted to function in an inflationary environment.

It is generally believed that stabilization programs aimed at stopping inflation involve an initial cost in loss of output because of rigidities in past nominal contracts (as in Fischer 1988 and Taylor 1979) or because of credibility problems. The costs of disinflation are therefore borne in the early stages of stabilization and fall later as the link with the past is severed and credibility is restored. This classical scenario of an initial recession followed by resumption of normal activity is indeed observed when low or moderate inflations are stopped. The best known recent examples from industrial countries are the stabilization policies in the United Kingdom and the United States in the early 1980s, whereas among developing countries Costa Rica (1982–83) and the Philippines (1983–84) had two recent programs. (For other examples, see Cline and Weintraub

Miguel A. Kiguel and Nissan Liviatan are both with the Country Economics Department at the World Bank. The authors are grateful to Michael Bruno, Guillermo Calvo, Vittorio Corbo, Allan Drazen, and Carlos Vegh for useful comments and discussions and thank Jariya Charoenwattana for research assistance.

1981.) In all of these episodes inflation fell gradually but persistently, and once the objective of reducing inflation was achieved, economic activity recovered, usually reaching prestabilization levels.

The classical recessionary scenario is also observed in chronic inflation economies that implemented orthodox stabilization programs, which used money as the nominal anchor (money-based stabilizations or MBS). A good example is the monetarist Chilean program of 1974–75, in which tightened monetary and fiscal policies led to a deep recession. Likewise attempts in 1990 to bring down inflation in Argentina, Brazil, and Peru through a monetary crunch led to a fall in economic activity. These programs were not very successful in achieving significant and sustainable reductions in inflation.

The experience with stopping hyperinflations is different. As Sargent (1982), Dornbusch and Fischer (1986), and Sachs (1986) among others document, hyperinflations have been stopped quickly and with relatively small costs. These were orthodox stabilization programs, which relied on tight fiscal policy; stabilization of, or in some cases fixing, the exchange rate; and no use of price controls. Although new evidence suggests that these programs were recessionary (for example, Wicker 1986), the induced recessions were less severe than those resulting from programs aimed at eliminating moderate inflation.

In this study we show that the relation between disinflation and output assumes an entirely different form in chronic inflation countries that used the *exchange rate* as the nominal anchor (we shall refer to it as exchange rate-based stabilization or ERBS). These countries experienced a small (or no) recession initially, quickly followed by reductions in inflation, which were accompanied by expanding output above the historical trend and falling unemployment. The expansionary phase sometimes lasted for several years before ending in a recession. In most of these programs inflation was reduced gradually, and hence the programs did not achieve the spectacular results achieved in ending hyperinflations. The main exceptions were the heterodox programs, in which rapid reductions in inflation were achieved by combining the use of the exchange rate as the nominal anchor with price and wage controls.

Given the differences in the outcomes of ERBS programs aimed at stopping hyperinflations and chronic inflation, it is difficult to argue that the effect on output is primarily due to the exchange rate policy as such. The view taken here is that the business cycle observed in ERBS in chronic inflation countries results from the combination of using the exchange rate as the anchor and the lack of credibility in the program's success. The perception of the program as temporary induces a boom in demand in anticipation of a collapse, which raises the level of economic activity following implementation of the ERBS. It will be argued that an expected collapse of a MBS does not generate the same expansionary response.

There are several differences between hyperinflationary and chronic inflation economies, which lead to differences in the credibility of disinflation policy. In hyperinflations the stabilization program is credible because the inflation pro-

cess is explosive and agents realize that there is no alternative to immediate stabilization. In addition in hyperinflations it is easy to identify the change of policy regime, which occurs when the money financing of extremely large fiscal deficits is ended. Finally, hyperinflation in Europe and Bolivia occurred in traditionally low-inflation economies where price stability was considered the normal state of affairs, which made stabilization more credible. Chronic inflation economies, in contrast, have developed a highly sophisticated ability to live with inflation, which makes stabilization postponable and therefore less credible. Also accelerations of inflation in these economies are more difficult to diagnose; in particular they are quite often unrelated (directly) to fiscal causes. Uncertainty about the causes of inflation tends to reduce credibility in stabilization policies. Finally, a long tradition of failed stabilizations in most chronic inflation economies makes any new attempt rather dubious. (These considerations are discussed in greater detail in Kiguel and Liviatan 1988.)

The article is organized as follows. Section I develops a simple analytical framework that shows that MBS and ERBS programs are likely to have different outcomes. After briefly examining MBS programs, section II discusses the main stylized facts concerning the business cycles in the ERBS. Section III elaborates on the reasons for the differences in the outcomes of ERBSs and MBSs, and section IV concludes with some policy issues.

I. Differences between Money-Based and Exchange Rate-Based Stabilizations

The possibility that disinflation programs that use money as the nominal anchor lead to different outcomes than those that use the exchange rate was raised by Rodriguez (1982, 1984), Fischer (1986), Helpman and Razin (1987), and Calvo and Vegh (1990) among others. Rodriguez (1984) develops a useful framework that produces the type of asymmetry in output cycles observed in the programs analyzed here. As in the work of Fischer, the results largely depend on the existence of inflationary rigidities (or price stickiness). The basic model assumes two goods: tradables and nontradables. The price of tradables is determined according to purchasing power parity, whereas the price inflation of nontradables is determined by expectations of future inflation and the excess demand prevailing in that market. Rigidities in the inflation process are introduced through the assumption of adaptive expectations. Finally, it is assumed that there is perfect capital mobility (in the sense that the interest rate parity condition holds), whereas the expected rate of depreciation is determined by the expected rate of inflation in nontradables and by the difference between the actual and the long-run equilibrium real exchange rate.

Under these conditions the real interest rate, r, is given by

$$r = i - \pi^e = i^* + \beta(e^e - \pi^e_N) + k \qquad (1)$$

where i is the nominal domestic interest rate, π^e is the expected rate of inflation,

i^* is the foreign interest rate, β is the weight of nontradables in the price index, ϵ^e is the expected rate of devaluation, π_N^e is the expected rate of inflation in nontradables, and k is a constant risk factor. When the official rate of devaluation (ϵ) is reduced, expectations on the devaluation rate will follow closely if the policy is credible. However, when inflationary expectations on nontradables exhibit downward rigidity (relative to ϵ), the fall in the expected devaluation rate will depress the real interest rate, thereby stimulating demand.

Within this framework Rodriguez (1984) shows that a disinflation policy based on reducing the rate of growth of the money supply results in a real appreciation, increased real interest rates, reduced domestic absorption (that is, recessionary pressure), and an improved trade balance. Alternatively, a disinflation program based on reducing the rate of devaluation of the exchange rate, as in Rodriguez (1982), results in a reduced real interest rate and an appreciated real exchange rate (the latter effect resulting from price stickiness). Although the impact on aggregate excess demand is in principle ambiguous (because the changes in the real exchange rate and the real interest rate exert pressures in opposite directions), Rodriguez shows that initially the interest rate effect dominates. Eventually, the real appreciation effect dominates and eliminates excess demand. If the Rodriguez model is modified appropriately, the cycle in demand will be associated with corresponding cycles in output and the trade balance.

In Fischer (1986, 1988), where price stickiness results from staggered, long-term nominal wage contracts in a setting of rational expectations, the author arrives at conclusions similar to those in Rodriguez (1984) with respect to the recessionary effects of MBS programs. He also points out that ERBS programs can have an initial expansionary effect because of the falling real interest rate. On the basis of numerical simulations, however, he concludes that in ERBS programs the more likely outcome is the recessionary scenario.

As will be argued later, agents' perception of a stabilization program as temporary (the credibility problem) can introduce a difference between the output patterns in the two policies. This difference is likely to arise if agents shift expenditures to the present in anticipation of a failure of the ERBS that will be associated with a balance of payments crisis. However, a failure of a MBS does not lead to intertemporal expenditure switching because the central bank does not commit its reserves to support the exchange rate (so there is no anticipation of a balance of payments crisis).

These arguments provide a basis for expectations of an asymmetry in the adjustment of the economy under monetary and exchange rate rules. These issues will be discussed in more detail after the empirical evidence on stabilization programs under both rules has been reviewed.

II. Empirical Evidence on the Business Cycle in Exchange Rate-Based Stabilizations

This section presents the stylized features associated with ERBS. As a background the outcomes associated with MBS are discussed.

Money-Based Stabilizations

The outcomes of MBS programs are well documented for industrial countries. Two recent programs are those of the United States under Volcker (see Dornbusch and Fischer 1987) and the United Kingdom under Thatcher (see Sargent 1986). Spain implemented a lesser known money-based program in the late 1970s. Examples of money-based programs in chronic inflation countries that will be referenced here are the 1958 program in Argentina under Frondizi, the Chilean program of 1974–75, the initial phase of the Argentine program under Martinez de Hoz in 1976–77 (all described in Kiguel and Liviatan 1988), and three programs launched in 1990 in Argentina, Brazil, and Peru aimed at stopping hyperinflation.[1]

The monetarist phase of these programs consisted of a sharp tightening in monetary policy, although in most cases, and certainly in all of the Latin American experiences, the target rate of monetary growth was not explicitly announced. In two cases—Chile and Argentina in the mid-1970s—the monetary tightening was part of an initial stage of an anti-inflation process, and the programs later on switched to ERBS. This change in strategy has not yet materialized in the very recent stabilization attempts in Brazil and Peru but has already taken place in the Cavallo stabilization in Argentina in 1991.

As is well known, MBS programs are recessionary in industrial countries. A tightening of monetary policy in the United States and the United Kingdom in the late 1970s and early 1980s brought down inflation, but at the cost of higher unemployment and lower output growth. It is not difficult to find other money-based programs in industrial countries with similar outcomes. The disinflation program in Spain in 1977–80, for example, also induced a recession; a sharp contraction in monetary growth led to a drop in inflation and growth in gross domestic product (GDP) as well as an increase in unemployment during the first two years of the program.

The recessionary effect of money-based programs is also observed in chronic inflation countries. For example, the fiscal MBS programs of Argentina of the late 1950s and the Argentine and Chilean programs of the mid-1970s had a clear recessionary effect on domestic output and employment (especially in Chile). In the more recent (1990) programs in Argentina, Brazil, and Peru, where tight money was a central part of the stabilization effort, the initial reduction in inflation (from hyperinflation levels) was associated with a deepening of the prevailing recessionary trend. In the programs in which the exchange rate was allowed to float, the recessionary impact was accompanied by an increase in real interest rates and a real appreciation as predicted by theory (in Brazil the real appreciation was observed in the parallel (free) exchange rate) and in most cases by an improvement in the trade balance. Chile of the mid-1970s is the one case where the MBS did not lead to real appreciation, although this is perhaps ex-

1. The recent programs in Argentina, Brazil, and Peru are not yet discussed in published form, but

plained by the fact that the exchange rate was sharply devalued to offset a severe deterioration in the terms of trade.

Exchange Rate-Based Stabilizations

This study concentrates on ERBS programs in chronic inflation countries in Latin America—namely Argentina, Brazil, Chile, Uruguay and Mexico—but includes also Israel, whose economy falls into the chronic inflation category. Each of these countries has pursued many stabilization programs, but only the 12 "major" ones, that is, those in which the public could recognize new initiatives that constituted a drastic break with previous policies, are included.[2] (See table 1 for program characteristics). A common feature of these programs is that each had major effects on the economy (for better or worse) and brought about significant reductions in inflation. In most cases the programs failed to stabilize inflation over a long time span (in which cases inflation accelerated), but in some instances the programs were part of a longer-term stabilization effort during which inflation was kept at a low level.

To gain an historic perspective on stabilization efforts, the empirical analysis covers programs implemented during the past three decades. In general the stabilization programs of each decade shared important common elements concerning the diagnosis of the causes of the inflation and the design of the appropriate policies to deal with it. For example, the stabilization programs of the 1980s—which include the Austral plan in Argentina, the Cruzado plan in Brazil, the Israeli stabilization of 1985, and the Mexican Pacto of 1988—were all heterodox programs relying on income policies. By contrast the Southern Cone stabilizations, implemented by the military governments in Argentina, Chile, and Uruguay in the 1970s (known as the Tablitas) were orthodox programs with a free market approach and an emphasis on liberalizing foreign trade and capital flows as part of the stabilization process. The 12 programs in the study contain several versions of exchange rate management as part of a disinflation policy, including fixed exchange rates and crawling pegs with various degrees of capital mobility (see table 1 and Kiguel and Liviatan [1990] for a fuller discussion).

In discussing the effects of ERBS on output, the balance of payments, relative prices, consumption, and investment, only the programs in Chile and Israel are presented in detail. These two representative programs share some of the important features observed in the other programs. (The complete set of figures for all the programs is presented in Kiguel and Liviatan 1990.) The basic outcomes for

2. These programs have been analyzed in numerous articles and books. The reader is referred to de Pablo and Martinez (1988) and Di Tella and Dornbusch (1989) on the Argentine stabilization programs, Di Tella (1979) on the Peronist program of 1974, Thorp (1965) on the Frondizi program of 1960, de Pablo (1974) on the Krieger-Vasena program of 1967, and Heymann (1987) and Kiguel (1991) on the Austral plan of 1985. For Chile's stabilization in the 1970s see Corbo (1985), Edwards and Edwards (1987), and Foxley (1980). For Brazil in 1964 see Simonsen (1974), and for the Cruzado in 1986 see Modiano (1988). For Uruguay in 1968 see Viana (1988) and Finch (1979), and for Mexico in 1988 see Ortiz (1990). *For Israel in 1985 see Bruno and Piterman (1988).*

all the programs are summarized in table 2, and figures 1–4 show more detailed results for Chile and Israel.

Real activity. In general economic expansion began soon after the stabilization programs were initiated. In Chile, where the use of the exchange rate for stabilization purposes began as early as the second semester of 1976, the whole period of ERBS up until 1982 was one of uninterrupted GDP growth. In Argentina the growth of output is evident in all five stabilizations. The upsurge of growth is also apparent in Uruguay after stabilization of the exchange rate in 1979. The more recent ERBS programs generally yielded similar results on output. Israel enjoyed high rates of growth in the business sector in the first three years of the programs; similar, although shorter, growth spans were observed in the Austral and Cruzado plans. The behavior of the unemployment rate was usually compatible with GDP growth, that is, unemployment fell in the growth phase of the cycle.

There is a slight difference between output behavior under orthodox and heterodox programs in the initial stage of the ERBS. In the orthodox programs the exchange rate policy was introduced when the inflation rate was already declining, having been dealt with initially by a monetary-fiscal package as in the Frondizi and Southern Cone stabilizations of the 1970s. In these cases there are no recessionary effects evident with the shift to a policy of reducing the rate of devaluation. In the heterodox stabilizations, however, exchange rate control was introduced along with income policies to stop the inflationary acceleration. In fact the income policies of the heterodox programs can be thought of as the counterpart to the monetary measures that preceded the ERBS in the orthodox programs of the Southern Cone. The fiscal adjustments induced an initial recessionary effect into those heterodox programs that undertook them, but the recessionary effect was both small and short-lived. When the period of exchange rate stabilization extended over a considerable time, as in the Southern Cone stabilizations of the 1970s, in Uruguay 1969, and in Israel, the recessionary phase began before the large maxi-devaluations set in.

In order to better gauge the cycle, figure 5 presents deviations of per capita income from the long-term trend. The trend was computed by fitting a linear or quadratic equation for log GDP per capita in a piecewise manner using appropriate intervals. (For Argentina the periods used to calculate the trend overlap, so the residuals from the overlapping parts were averaged.) These diagrams confirm that in 11 out of 12 cases there emerges an expansionary phase of output relative to trend in the course of the ERBS (the exception is the Brazilian stabilization of 1964–67, which was characterized by a continuous recession relative to trend). Moreover, except for the Krieger Vasena stabilization of 1967–70 in Argentina, all the expansions (relative to trend) begin around the time the exchange rate starts its role as the nominal anchor. In these respects the expansions in chronic inflation countries differ from those sometimes observed in post-hyperinflations, as in Germany, where industrial output remained well below

(*Text continues on page 296.*)

Table 1. *Description of Exchange Rate-Based Stabilization Programs*

Country and program dates (year, quarter)	*Exchange rate*	*Incomes policies*[a]	*Fiscal adjustment*	*Reduction in monthly inflation*		*Commercial policy reform*	*Preceded by corrective monetary or fiscal measures*	*Preceded by maxi-devaluation*
				From	*To*			
Israel 1985.1–present	Fixed with infrequent adjustment	Yes (shock)	Large initial adjustment	21.2	6.1	No	No	Yes
Argentina (Austral) 1985.1–1986.3	Fixed	Yes shock	Large transitory adjustment	24.9	2.6	No	No	Yes
Brazil (Cruzado) 1986.1–1986.4	Fixed	Yes (shock)	No	11.1	1.7	No	No	No
Mexico 1988.1–present	Fixed (first year) then crawl	Yes (small shock)	Main adjustment before program	8.2	2.6	Trade and capital account liberalization	Fiscal adjustment	Yes
Chile 1976.3–1982.3	Crawl, then preannounced then fixed	No	Yes	11.2	6.5	Trade and capital account liberalization	Yes	Yes
Uruguay 1978.4–1982.4	Preannounced	No	Yes	3.4	4.6	Liberalization	Trade and capital adjustment	No (small fiscal)

Country and program dates (year, quarter)	*Exchange rate*	*Incomes policies*[a]	*Fiscal adjustment*	*Reduction in monthly inflation*		*Commercial policy reform*	*Preceded by corrective monetary or fiscal measures*	*Preceded by maxi-devaluation*
				From	*To*			
Argentina 1978.4–1981.1	Preannounced	No	Moderate adjustment	8.1	8.6	Trade and capital account liberalization	Yes	No
Argentina 1973.3–1975.2	Fixed	Yes	No	5.7	0.8	No	No	No
Argentina 1967.2–1970.3	Fixed	Yes (gradual)	Yes	2.5	2.7	Incentives for capital inflows	No	Yes
Brazil 1964.2–1968.3	Fixed with step devaluation	Yes (gradual)	Yes	6.4	4.2	No	No	Yes
Uruguay 1968.2–1972.1	Fixed	Yes (shock)	Initially yes, deterioration later on	9.5	1.9	No	Yes (6 months before)	Yes
Argentina 1959.3–1962.2	Fixed	No	Initially yes, deterioration later on	9.5	1.9	Incentives for foreign investment	Yes (IMF program 6 months before)	Yes

Note: For a graphic trend of the performance of these indicators during the program, see Kiguel and Liviatan (1990).

a. "Shock" means a drastic and immediate reduction in inflation—as in a price freeze. "Gradual"—incomes policies implemented in small steps, spread over time—as in gradual adjustments of staggered wage contracts.

Table 2. *Outcomes of Exchange Rate-Based Stabilization Programs*

Country and program dates (year, quarter)	*Year in which upswing in GDP growth relative to trend began*[a]	*Current account*	*Was there a consumption boom?*	*Was there an investment boom?*	*Real exchange rate*	*Fiscal deficit/GDP*	*Real wages*
Israel 1985.1–present	First	Improves until 1987, then deteriorates	Yes	Yes, initially	Gradually appreciates	Decreases	Increases
Argentina (Austral) 1985.1–1986.3	First	Deteriorates	Yes	Yes	Appreciates	Decreases initially, then increases	Decreases
Brazil (Cruzado) 1986.1–1986.4	First	Deteriorates	Yes	Unclear	Appreciates	No change	Increases
Mexico 1988.1–present	Second	Deteriorates	Maybe	Yes	Appreciates	Decreases	No change
Chile 1976.3–1982.3	First	Deteriorates	Yes	Yes	Depreciates until 1979, then appreciates until 1981, then depreciates	Decreases	Increases until 1982.1, then decreases
Uruguay 1978.4–1982.4	First	Deteriorates	Yes	No	Appreciates	Increases	Decreases
Argentina 1978.4–1981.1	First	Deteriorates	Yes	Yes	Appreciates	Increases	Initially increases, then decreases

Country and program dates (year, quarter)	Year in which upswing in GDP growth relative to trend began[a]	Current account	Was there a consumption boom?	Was there an investment boom?	Real exchange rate	Fiscal deficit/GDP	Real wages
Argentina 1973.3–1975.2	First	Initially improves until 1974.1, then deteriorates	Yes	No	Appreciates	Increases	Increases until 1974, then decreases
Argentina 1967.2–1970.3	Second	Deteriorates	Yes	Yes	Appreciates	Decreases until 1969, then increases	Increases
Brazil 1964.2–1968.3	Fourth	Improves until 1965, then deteriorates	Not clear	Yes	Appreciates until 1967.1, then depreciates	Decreases	Decreases until 1967.1, then increases
Uruguay 1968.2–1972.1	First	Deteriorates	Yes	Yes	Appreciates, then depreciates	Decreases until 1970, then increases	Increases until 1971.1, then decreases
Argentina 1959.3–1962.2	First	Deteriorates	Yes	Yes	Appreciates	Decreases 1960, then increases	Increases

Note: For a graphic trend of the performance of these indicators during the program, see Kiguel and Liviatan (1990).
a. Refers to number of years after beginning of program.

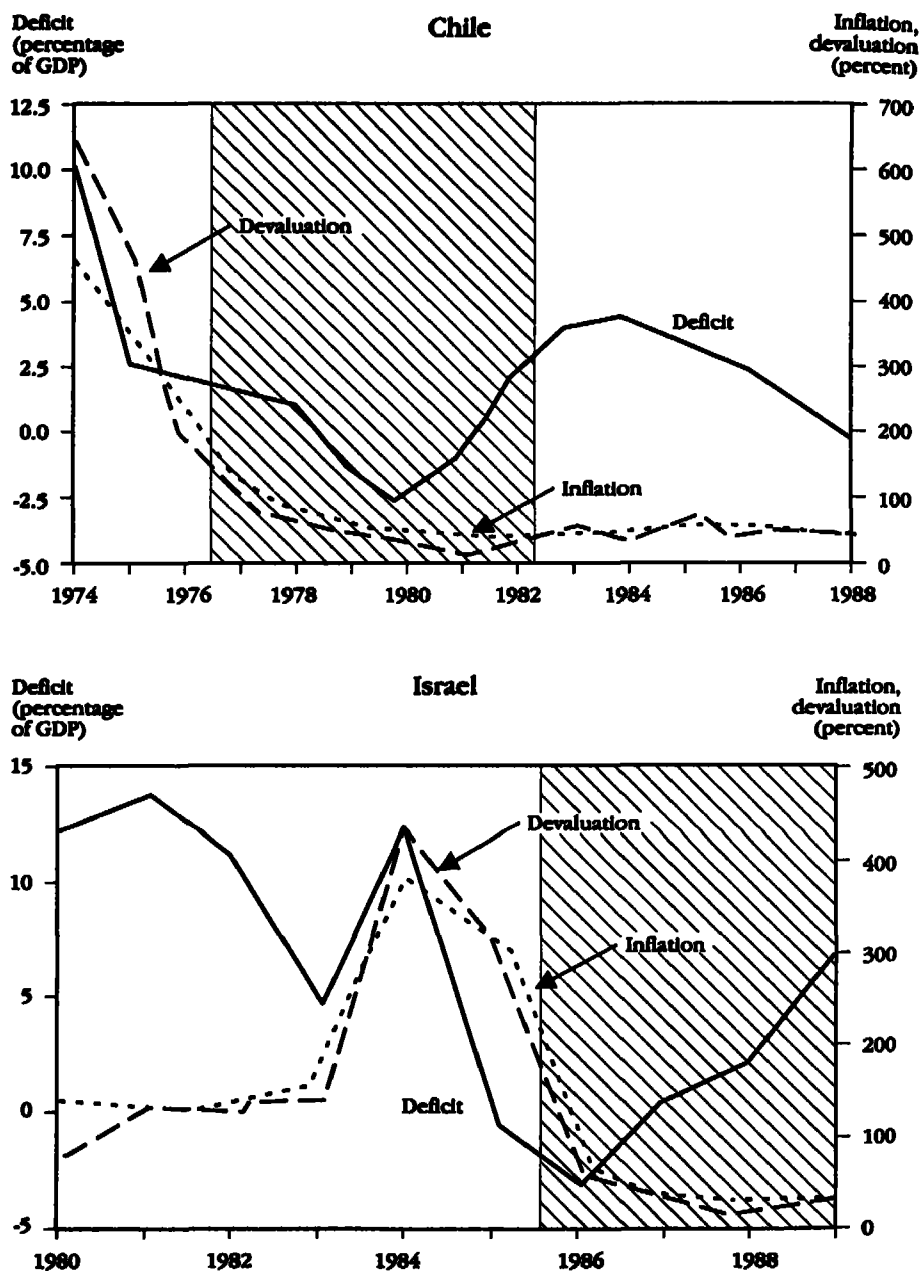
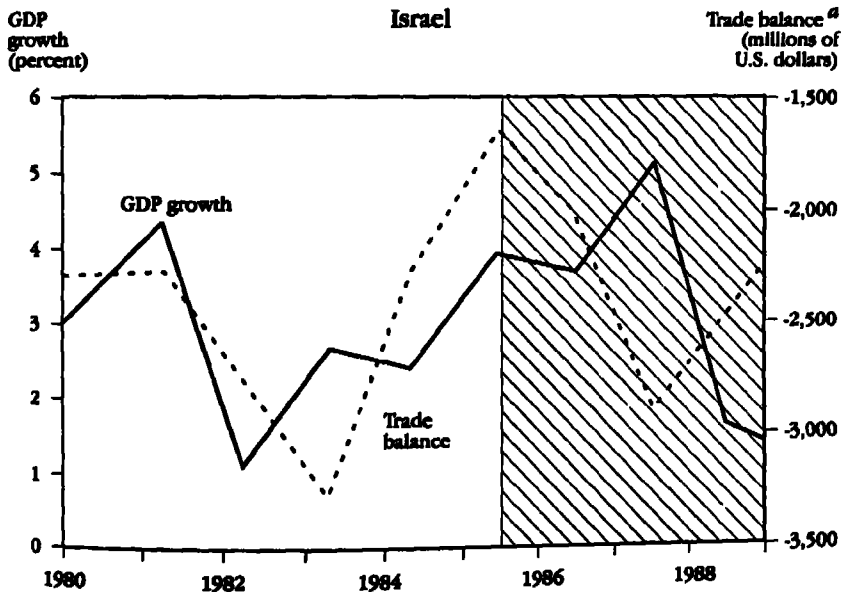
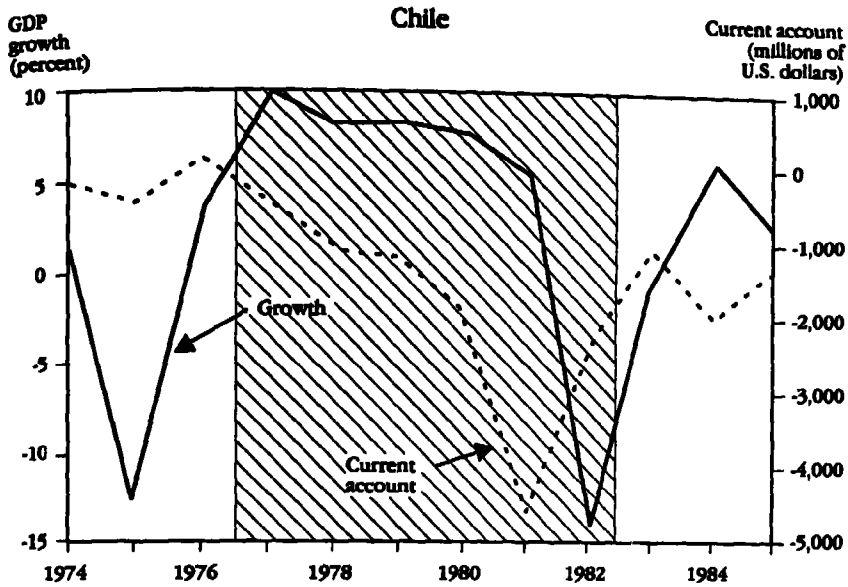
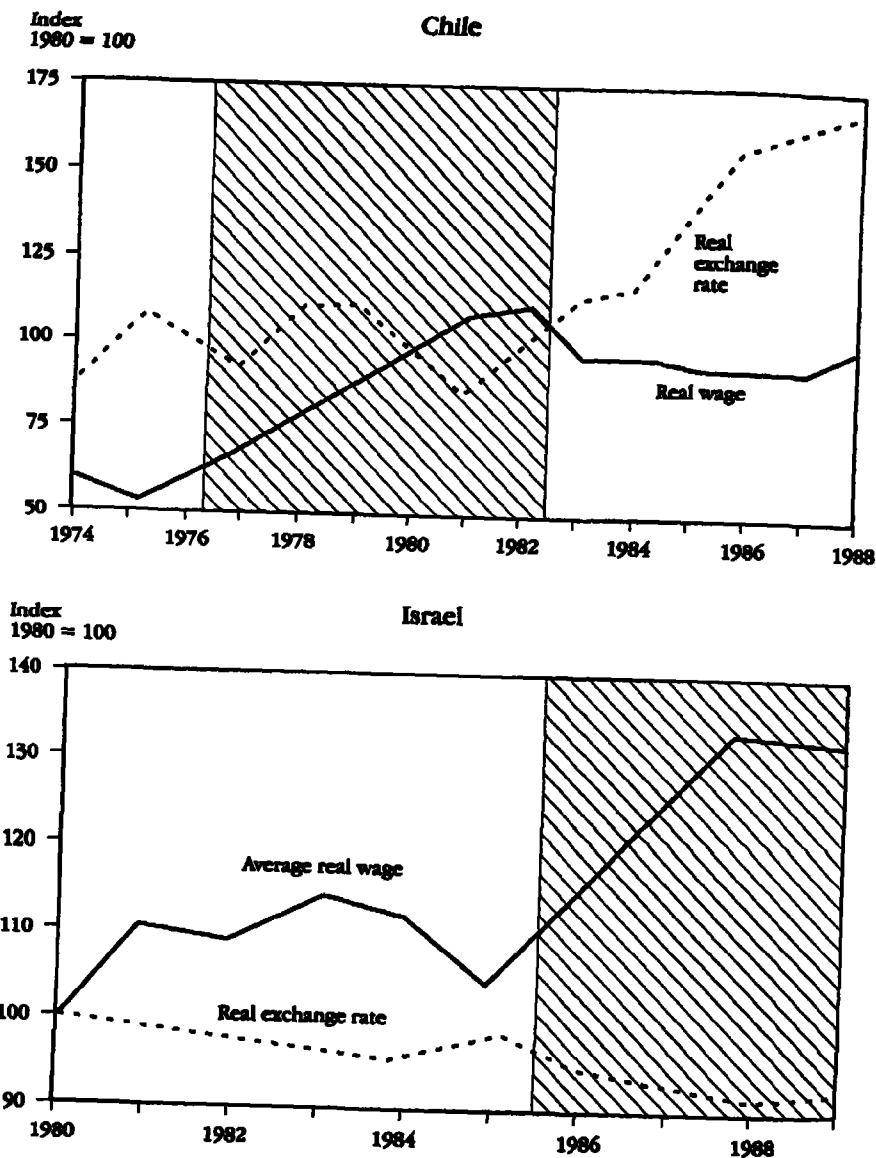
Figure 1. *The Fiscal Deficit, Inflation, and Devaluation in Chile and Israel*

Source: For Chile, Banco Central de Chile and INE; for Israel, Bank of Israel.

Figure 2. *GDP Growth and the Current Account*

a We use the trade balance rather than the current account balance because the latter contains a large component in the form of U.S. grants.

Source: For Chile, Banco Central de Chile; for Israel, Bank of Israel.

Figure 3. Real Exchange Rate and Real Wage in Chile and Israel

Note: The real exchange rate is the average of real prices of imports and exports.

Source: For Chile, Banco Central de Chile and World Bank (various years); for Israel, Bank of Israel (various years).

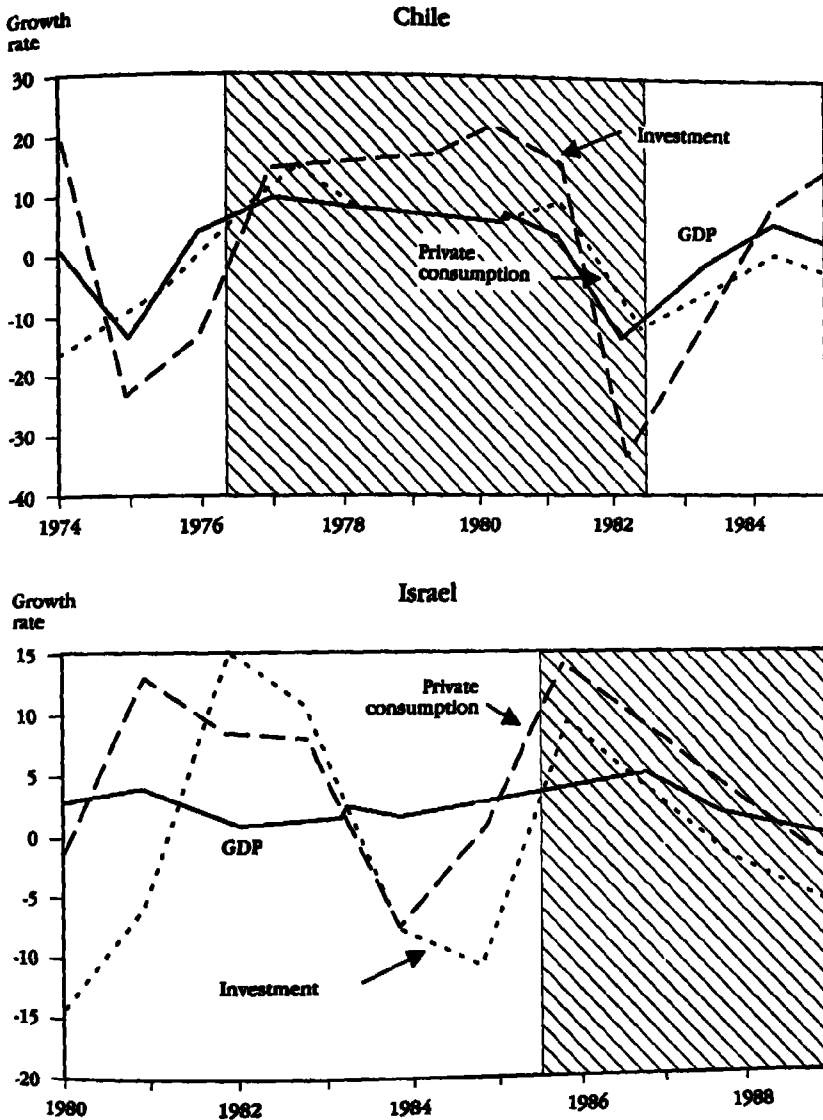
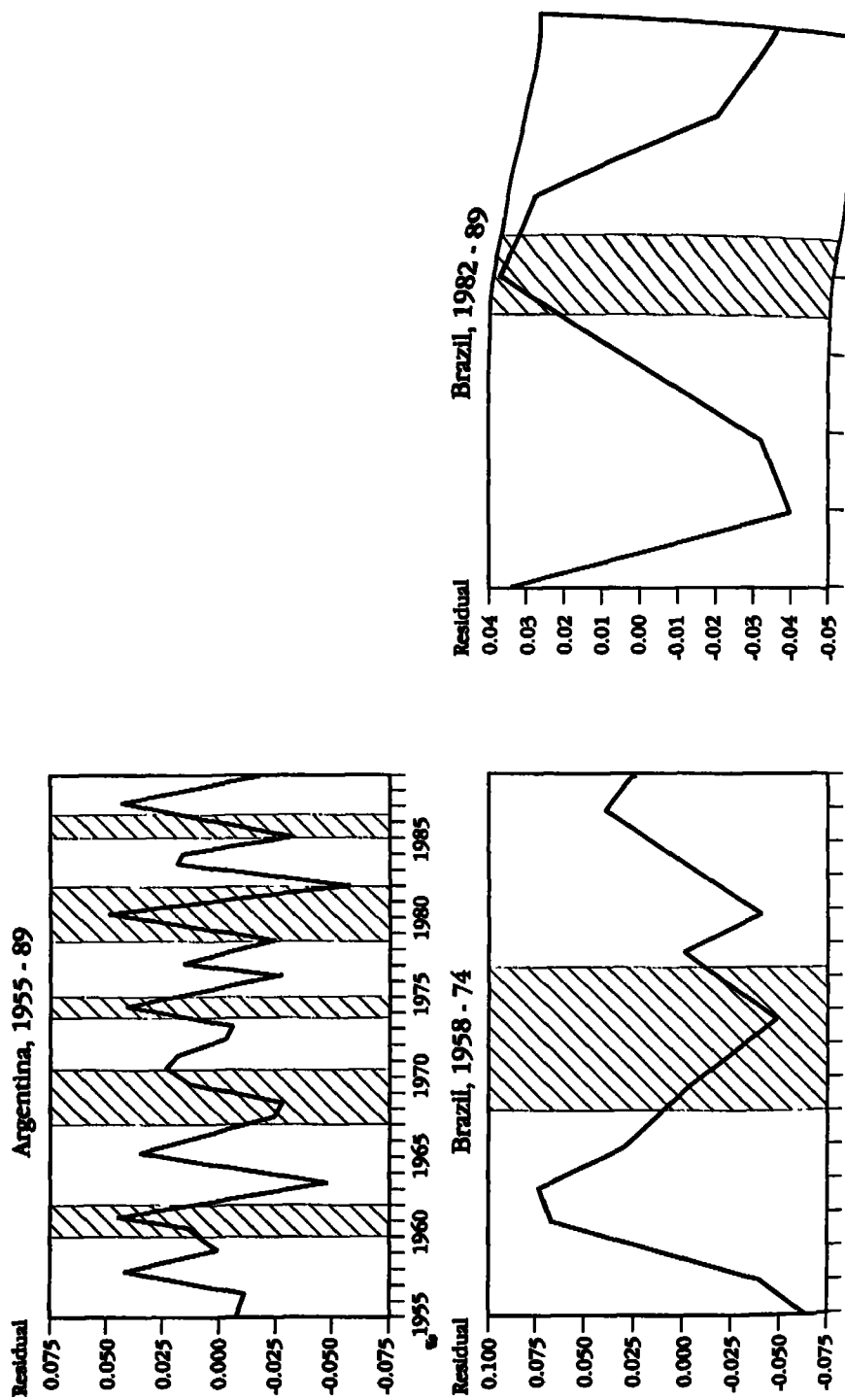
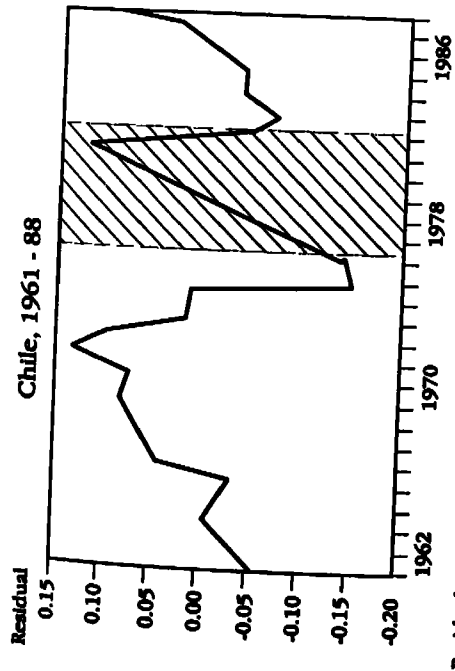
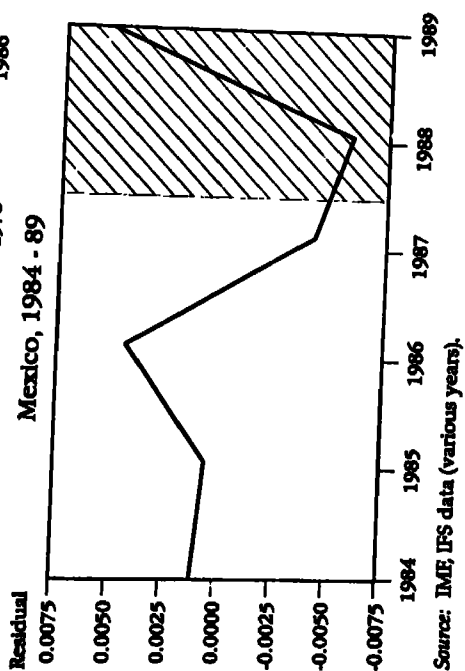
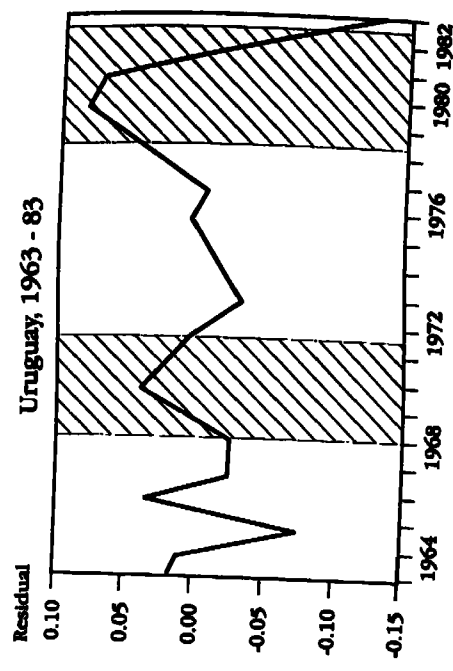
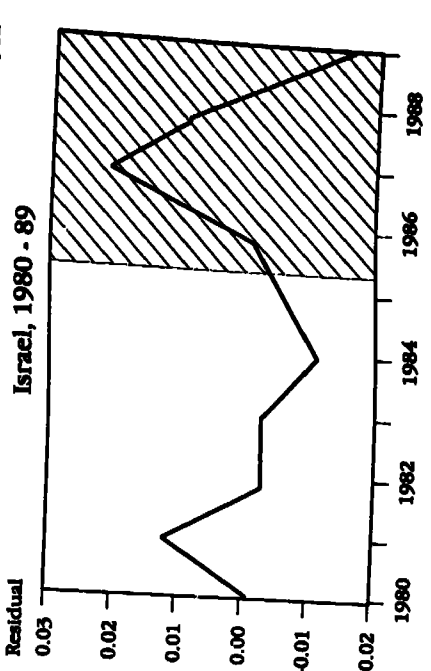
Figure 4. *Real Rates of Growth of GDP, Private Consumption, and Investment in Chile and Israel*
(percent)

Source: For Chile, Banco Central de Chile; for Israel, Bank of Israel.

Figure 5. *Deviations of Log GDP per Capita from Trend*

Source: IMF, IFS data (various years).

trend (Garber 1982). Also most of the ERBS programs were preceded by recessionary period relative to trend, which indicated the existence of excess capacity that may have provided suitable conditions, from the supply side, for the upswing.

The balance of payments. All ERBS programs were associated with a deterioration of the trade balance and the current account during the program. The expansionary force of these programs, from the point of view of domestic uses, was therefore even more pronounced than with the output growth criterion. During the expansionary phase, the current account typically deteriorates (as illustrated in figure 2). The capital inflows that financed these deficits were, as a rule, reversed at some later stage of the program, thus coinciding with the beginning of the recessionary phase. The inability to finance the growing current account deficits was, in most cases, the immediate reason for the end of the boom. Two important exceptions are Brazil in 1964–67, in which capital inflows continued to finance the current account deficit and made growth sustainable for many years to come, and Israel, in which the boom ended without a marked deterioration of the current account.

Relative prices. As a rule, real wages increased with the upswing of economic activity (see figure 3), although in some cases this occurred with a lag, which may be explained by the following. First, the real wage may have been raised up front in order to compensate for its anticipated erosion if the reduction in inflation was sluggish. A temporary reduction in the real wage in the early phase of stabilization would then be observed, as in the Krieger Vasena stabilization. Second, the real wage may have been deliberately kept below its equilibrium level for some time by income policies, as in Israel. The normal behavior of the real exchange rate during the boom was (as expected) in the opposite direction to that of the real wage. During the periods of a full peg, the real exchange rate fell especially quickly.

Consumption and investment. Most of the expansions in output during ERBS were accompanied by a consumption boom (see table 2). Clearly, when GDP grows faster, consumption is expected to grow faster also, although the growth in consumption should lag behind the growth in output if the stimulus to GDP did not originate from the consumption side. The term "consumption boom" refers to the case in which consumption grows faster than GDP when growth in the latter accelerates or is above normal. The most conspicuous examples of a consumption boom took place in the Peronist and Martinez de Hoz stabilizations in Argentina, Uruguay in 1969, and the Israeli program (shown in figure 4).

Increased investment played a dominant role in the expansions during the Argentine programs of the 1960s, much of it induced by government policies. In the programs of the 1970s the Chilean ERBS was driven by a continual investment boom. Although investment booms were also characteristic of the pro-

grams of the 1980s, these booms were not as large as in earlier programs (with the possible exception of Mexico). This was partly related to reductions in capital inflows and to increases in macroeconomic instability following the debt crisis of the 1980s. In Israel a short-lived upsurge in investment occurred, but the ratio of investment to GDP was lower after the stabilization than before.

The fiscal deficit. The expansion in output following implementation of the ERBS occurred in spite of sharp reductions in fiscal deficits. This is most evident in Israel, where the elimination of the fiscal deficit in 1986 coincided with a sharp consumption boom. This boom is quite surprising because the increase in taxation clearly outweighed any possible reduction in the inflation tax. A similar phenomenon occurred in Chile, where the fiscal deficit was turned into a surplus during the ERBS. And the cut in the fiscal deficit before, or along with, the expansion in aggregate demand during the early phase of the ERBS is also characteristic of all Argentine programs except the Peronist stabilization, in which the deficit increased from the start.

In most stabilizations the initial reduction in the fiscal deficit was later reversed. In the Argentine programs, in particular, the loosening of fiscal policy undermined the disinflationary exchange rate policy. The systematic nature of the fiscal reversals lends support to the view that agents might have treated the stabilizations as temporary measures, which is important for understanding the nature of the cycle. In some cases, notably in Chile and Israel, the poststabilization increases in the fiscal deficit were largely endogenous to the recessionary phase of the business cycle. In these cases the reversal in the fiscal deficit was regarded as temporary; old inflationary expectations were not therefore rekindled.

III. Theoretical Aspects of the ERBS Business Cycle

One cannot expect to obtain any useful insight into the cyclical behavior of ERBS from the classical models of the representative individual in an economy with flexible prices and fully credible policies. Indeed it has been shown by Obstfeld (1985) and Calvo (1986) that, in the foregoing framework, a permanent reduction in the constant rate of devaluation is entirely neutral. The same conclusion holds for MBS (see discussion in Kiguel and Liviatan 1990). In order to explain the boom in output and demands (and the subsequent recession), as well as the behavior of the current account and relative prices during ERBS programs, considerations of credibility and of price stickiness must be introduced.

In section I, following Rodriguez and Fischer, it was shown that since a reduction in the real interest rate stimulates aggregate demand, an ERBS implemented in the presence of inflationary rigidities or price stickiness can generate an expansionary phase. To what extent is this real interest rate factor relevant for the explanation of the actual cycles in ERBS programs? It seems that in the Southern Cone stabilizations in the late 1970s the reduction in the real interest

rates was quite evident. Thus Corbo (1985) reports that the Tablita policy Chile, and the increased capital inflows with which it was associated, led t downward pressure on domestic interest rates, which stimulated a rise in aggregate demand. Similar findings are reported in Ramos (1986) for the early stages of all the Tablita policies for lending rates. It is also interesting that in all Southern Cone Tablitas of the late 1970s, recession along with high real interest rates began before the collapse of the exchange rate regime, as suggested by the modified Rodriguez model.

The fall in real interest rates in the early stages of the Tablitas is a property that was not shared by the stabilization programs of the 1980s, partly because the debt crisis limited capital flows. Real interest rates in Israel and Mexico rose to extremely high levels during stabilization, which may be related to the lower degree of credibility in these programs due to the sharp reduction of inflation with the aid of controls. The rise in the ex ante real rates was smaller but apparently still significant. In Israel an estimate of π^e computed from capital market data shows that the ex ante real interest rate on bank loans rose quite significantly in the early stage of stabilization (Bank of Israel 1985). It seems therefore that in these cases something more than the Rodriguez interest rate mechanism is needed to explain the expansion.

The foregoing models of Fischer and Rodriguez assume that the disinflation program is eventually successful. The fact that so many stabilization programs fail, however, must lead to pessimistic views about the chances of any new program to succeed. The very expectation that the stabilization is only temporary may give rise, in the early stages, to an expansion of aggregate demand.

The issue of expectations that stabilization is temporary has been investigated in recent papers by Calvo (1986, 1987, 1991). These cash-in-advance models show that agents who expect stabilization to be temporary will shift part of their future consumption expenditures to the present. In the present, when the rates of devaluation and inflation are low, the cost of holding money (which is necessary to carry out purchases) is also low, whereas the opposite is true for future periods when high inflation is expected to resume. This gives rise to increased expenditures in the stabilization period, accompanied by current account deficits and real appreciation. All these features are clearly consistent with the phenomenon of the consumption boom and the related developments described earlier.

Calvo's model is a Ricardian one, with the property that a permanent reduction in the (constant) rate of devaluation is neutral. This underscores the fact that the initial consumption boom is related entirely to expectations of temporariness. Unlike the Rodriguez and Fischer models, where the boom is the result of an initial reduction in the real interest rate, the rise in consumption in Calvo's model is caused by a temporary reduction in the nominal interest rate. The level of the real interest rate, which is constant in Calvo's model, has nothing to do with the cyclical behavior of demands. This helps to explain the emergence of the consumption boom in programs where the real interest rate was very high, as in Israel.

The cash-in-advance setting, which stresses the role of liquidity, seems (implicitly) to be related more closely to a consumption boom in terms of durables. Indeed the data for Israel show that durable purchases increased tremendously in 1986, (up 47 percent from the previous year) after the ERBS program was implemented in July 1985.

Somewhat paradoxically, one can arrive at similar conclusions about the consumption boom if one assumes that consumers view the stabilization as permanent and that the reduced uncertainty about relative prices and about government policies enhances productivity. This sort of expectation may raise agents' expected permanent income and thus raise consumption. This theory, however, cannot explain the cyclical nature of the consumption boom unless it is assumed that in each case the expectations were incorrect. The explanation based on expected temporariness of the ERBS is more robust in the sense that it can explain the cyclical behavior of consumption regardless of whether agents' expectations turn out to be correct or incorrect (see Kiguel and Liviatan 1990 and Calvo and Vegh 1990).

An alternative model of the cycle, which is again based on expectations of temporariness (but not in a cash-in-advance setting) and on a boom of durables purchases was formulated recently by Drazen (1990). He shows that when the fixed exchange rate policy is expected to collapse at some definite date, there will be a run on imported durables just before the collapse. This may be a result of hedging against devaluation when there is no free access to foreign exchange or in anticipation of the imposition of quantitative restrictions following a balance of payments crisis. In this case the ERBS will not induce an initial boom. When the date of collapse is uncertain, however, the wave of durables purchases will be spread over the span of the stabilization program. This will cause a boom in domestic output when expenditures on durables are complementary to consumption of domestically produced goods. Drazen's argument may be extended to claim that when the collapse takes the form of a balance of payments crisis, agents will expect a sharp tightening of credit conditions after the crisis and consequently will be motivated to advance all kinds of expenditures to the present. Drazen's model also can explain the phenomenon of the recession in output starting before the collapse of the exchange rate regime; this will occur when the realization of the collapse happens to occur at a relatively distant date in the future.

Drazen provides data on booms in expenditures on durables during ERBS programs in Argentina, Chile, and Mexico, which supplement the data here from Israel and provide more evidence supporting the theory that the motivation to advance purchases on durables is one of the important driving forces of the business cycle in an ERBS.

Since real wages tend to rise with the expenditure cycle, it is conceivable that the former may actually cause the latter because of the higher propensity to spend out of wage income. To be an adequate explanation, however, it has to be shown, first, that real wages affect demand more strongly than supply and, second, that real wages should rise more in an ERBS than otherwise.

The answer to the first question is certainly unclear from the theoretical point of view (see Krugman and Taylor 1978 and Lizondo and Montiel 1989). As for the second one, it is often mentioned (especially for the Chilean Tablita) that lagged wage indexation will cause real wages to rise when inflation is falling. However, this explanation does not work for the heterodox programs in which the foregoing type of inertia is eliminated at the outset.

Apart from the effects of disinflation on demand, there may be expansionary effects originating from the supply side. The strong effect of exchange rate stabilization on prices, which we often find in an ERBS, may increase efficiency by reducing excessive variation in relative prices and by shifting resources out of excessive financial and speculative activities. However, the latter effect operates only in the longer run. In the medium run the effect of restructuring the economy toward a low inflation environment may well be recessionary (see Garber [1982] for an example from the German stabilization).

The real interest rate effect, which is expansionary in the Rodriguez-Fischer models for an ERBS, works in an opposite manner in a MBS. With wage or price stickiness a monetary crunch will raise the real interest rate. (With perfect capital mobility this entails an initial real appreciation and an expected devaluation, as in Dornbusch's overshooting theory.)

If Calvo's (1986) model is adapted to a MBS, temporariness will still lead to a consumption boom. This boom requires an initial drop in prices; consequently, downward price rigidity may nullify the expansionary effect. Calvo and Vegh (1990) show that, in a model with staggered prices, a temporary ERBS is expansionary whereas a MBS is contractionary.

If one takes the view that expectations of a collapse associated with a balance of payments crisis and imposition of quantitative restrictions induce an advancement of purchases that generate a business cycle in an ERBS, then it is easy to see this factor does not operate in a MBS. In a MBS the central bank does not commit its foreign exchange reserves to protect the disinflation policy, and therefore the issue of the balance of payments crisis does not arise.

The forgoing analysis dealt with some general considerations that may explain expansionary tendencies in ERBS programs. In practice one may usually identify specific factors that contributed to the expansion of output in individual cycles. It is conceivable that these factors were no less important in generating the booms than the exchange rate policy itself. In some other cases the interaction of the exchange rate stabilization strategy with other factors created a mechanism for unsustainable expansion.

Most ERBS programs were initiated under favorable external conditions, which made it possible in some cases to pursue expansionary policies simultaneously with disinflation policies. In some cases, especially in the early stabilizations in Argentina, the pegging of the exchange rate was part of a broader development strategy that favored foreign investment and considered a stable exchange rate as part of the required financial environment. This approach may have resulted in both disinflation and expansion.

In some cases, as in the Tablita policies in the Southern Cone, the ERBS was part of a package that included the liberalization of the trade and capital accounts. Some authors stress the expansionary effect of the latter (Corbo 1985 and Edwards and Edwards 1987), but liberalization of the trade account can also be expansionary if it is perceived to be temporary.

IV. Policy Issues

The business cycle phenomenon associated with an ERBS does not in itself imply that this policy cannot be part of a longer-term successful process, as has been demonstrated in such countries as Chile and Israel. However, even in these cases the variation in economic activity and in consumption over time is undesirable for several reasons. First, a stable path of consumption will be preferred to a variable one when consumers have a concave utility function for consumption (see Calvo 1986). Second, excessive purchases of capital goods during the expansionary phase (in anticipation of a failure of the exchange rate policy) lead to an inefficient allocation of investment. Third, the difficulty in correcting both the overvaluation and the excessive real wages exacerbates the recessionary phase and leads to an unnecessary loss of output.

For most programs that turn out to be temporary, there are additional problems. In these cases the failure of an ERBS can destabilize the inflationary process, as was clearly the case in the aftermath of the Austral and Cruzado plans. The loss of credibility resulting from failure to stabilize implies that future stabilizations must employ even harsher policies, leading to more severe recession and loss of output.

This raises the question of whether the stabilization cycle is unavoidable in practice and, if not, what can be done to mitigate its effect? As for the first question, the evidence from the stabilizations of Brazil 1964–67 and Mexico 1988–89 show that the excessive expansionary phase can be avoided. This may have been possible in the Mexican case because fiscal adjustment was carried out long before 1988. In fact the fiscal deficit was drastically cut as early as 1983, and the budget even showed an operational surplus in 1987. Although the Mexican experiment is still very young and definite conclusions cannot yet be reached, it seems advisable to implement fiscal adjustment before stabilizing the exchange rate in order to enhance the credibility of the program. This, of course, may not be politically feasible; in Mexico this stepwise procedure was facilitated by the need to use recessionary fiscal measures to deal with the balance of payments crisis in the early 1980s.

Given the difficulties encountered in the ERBS, should policymakers prefer the MBS? Or can conditions be specified that can guide policymakers in choosing between the two strategies? It is certainly arguable that if the delayed recession cannot be avoided in the ERBS then it might be preferable to have it earlier by implementing a MBS. Formulation of this choice in terms of "recession now (MBS) or recession later (ERBS)" relates to the discussion in signaling theory (as in

Vicker 1986) of whether it is preferable to establish credibility up front by adopting a drastic policy (such as a MBS) or postponing the confrontation to a later date and accepting, for the time being, the cost associated with lack of credibility (as with the business cycle of the ERBS).

The conclusion from signaling theory (Kiguel and Liviatan 1988) is that a drastic policy may be preferable when initial credibility is low. In the current context this means that a MBS is preferable when the economy has a long history of failed stabilizations, as is quite common in Latin America. Indeed the recent preference for a strict MBS in Argentina and Brazil in 1990, after the failure of many stabilizations in the 1980s, seems to support this view.[3] However, when credibility is relatively high (that is, relative to the standards of chronic inflation countries), perhaps as a result of a basic improvement in the external position, an ERBS is preferable because the cost of not establishing initial full credibility in nominal anchors is relatively low. An ERBS was thus the appropriate policy choice in the case of Israel. The ERBS strategy has the additional advantage of making the fiscal adjustment transparent because of the relative stability of prices. And it is much easier for the public to monitor the exchange rate rule as compared with the complexity of monitoring some money supply anchor. Thus the policymaker may hope to alter the public's views on credibility during an ERBS by adhering to the exchange rate rule and by adopting a strict fiscal adjustment. Consequently, the recessionary phase may turn out to be relatively mild.

Another policy option is to adopt a stepwise disinflation strategy with an ERBS followed by a MBS. This may allow the policymakers to reap some of the gains of both policies. Chile, for example, did quite well after abandoning its fixed exchange rate regime in 1982—economic growth resumed jointly with a real devaluation and a stable, low level of inflation. Two difficulties arise with the proposed switch of anchors. First, if the excessive real appreciation during the ERBS is due to lack of credibility in the adherence to the nominal anchor, then this credibility issue will reappear in connection with the money supply rule as well. Second, adopting the money supply anchor is problematic if the economy remains highly indexed. With indexation, monetary shocks are translated very easily into price shocks, which may undermine credibility in price stability (given incomplete information about the source of the shocks). Chile overcame this difficulty by abolishing formal wage indexation, a decision that was facilitated by the balance of payments crisis of 1982. By contrast Israel remains highly indexed and so must continue with a policy of infrequent step devaluations, thus relying on the exchange rate as the nominal anchor and going through the ERBS business cycle. Its comfortable external position continues to support this reliance on the exchange rate as the anchor.

3. The shift to an ERBS in Argentina in 1991 took place after considerable gains in credibility of the disinflation process.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Banco Central de Chile. Various years. *Annual Report*. Santiago.

Bank of Israel. Various years. *Annual Report*. Jerusalem.

Bruno, Michael, and Sylvia Piterman. 1988. "Israel's Stabilization: A Two-Year Review." In Michael Bruno, Guido Di Tella, Rudiger Dornbusch, and Stanley Fischer, eds., *Inflation Stabilization*. Cambridge, Mass.: MIT Press.

Calvo, Guillermo A. 1986. "Temporary Stabilization: Predetermining Exchange Rates." *Journal of Political Economy* 94 (6): 1319–29.

———. 1987. "Balance of Payments Crises in a Cash-in-Advance Economy." *Journal of Money, Credit, and Banking* 19 (1): 19–32.

———. 1991. "Temporary Stabilization Policy: The Case of Flexible Prices and Exchange Rates." *Journal of Economic Dynamics and Control (Netherlands)* 15 (January): 197–213.

Calvo, Guillermo A., and Carlos A. Vegh. 1990. "Credibility and the Dynamics of Stabilization Policy: A Basic Framework." IMF Working Paper WP/90/110. Washington, D.C. Processed.

Corbo, Vittorio. 1985. "Reforms and Macroeconomic Adjustments in Chile during 1974–84." *World Development* 13 (8): 893–916.

Cline, William R., and Sidney Weintraub. 1981. *Economic Stabilization in Developing Countries*. Washington, D.C.: The Brookings Institution.

de Pablo, Juan Carlos. 1974. "Relative Prices, Income Distribution, and Stabilization Plans: The Argentine Experience, 1967–70." *Journal of Development Economics* 1 (3): 167–90.

de Pablo, Juan Carlos, and Alfonso Martinez. 1988. "Argentina." World Bank, Washington, D.C. Processed.

Di Tella, Guido. 1979. "The Economic Policies of Argentina's Labor-Based Government, 1973–76." In Rosemary Thorp and Laurence Whitehead, eds., *Inflation and Stabilization in Latin America*. New York: Holmes and Meier.

Di Tella, Guido, and Rudiger Dornbusch, eds. 1989. *The Political Economy of Argentina*. University of Pittsburgh Press.

Dornbusch, Rudiger, and Stanley Fischer. 1986. "Stopping Hyperinflation Past and Present." *Weltwirstchaftliches Archiv* 122 (1, January): 1–47.

———. 1987. *Macroeconomics*. 4th Edition. New York: McGraw Hill.

Drazen, Allan. 1990. "Can Exchange Rate Freezes Induce Business Cycles?" University of Maryland, Department of Economics, College Park. Processed.

Edwards, Sebastian, and Alejandra Cox Edwards. 1987. *Monetarism and Liberalization: The Chilean Experiment*. Cambridge, Mass.: Ballinger.

Finch, M. H. J. 1979. "Stabilization Policy in Uruguay since the 1950s." In Rosemary Thorp and Laurence Whitehead, eds., *Inflation and Stabilization in Latin America*. New York: Holmes and Meier.

Fischer, Stanley. 1986. *Indexing, Inflation, and Economic Policy*. Cambridge, Mass.: MIT Press.

———. 1988. "Real Balances, the Exchange Rate, and Indexation: Real Variables in Disinflation." *Quarterly Journal of Economics* 103 (1): 27–49.

Foxley, Alejandro. 1980. "Stabilization Policies and Stagflation: The Cases of Brazil and Chile." *World Development* 8: 887–912.

Garber, Peter M. 1982. "Transition from Inflation to Price Stability." *Carnegie-Rochester Conference Series on Public Policy* 16 :11–42.

Helpman, Elhanan, and Assaf Razin. 1987. "Exchange Rate Management: Intertemporal Tradeoffs." *American Economic Review* 77 (March): 107–23.

Heyman, Daniel. 1987. "The Austral Plan." *American Economic Review* 77 (2): 284–87.

Instituto Nacional de Estadisticas (Chile). Various issues. *Informativo estadistico*.

International Monetary Fund. Various years. *International Financial Statistics*. Washington, D.C.

Kiguel, Miguel A. 1991. "Inflation in Argentina: Stop and Go since the Austral Plan." *World Development* 19 (8): 969–86.

Kiguel, Miguel A., and Nissan Liviatan. 1988. "Inflationary Rigidities and Orthodox Stabilization Policies: Lessons from Latin America." *The World Bank Economic Review* 2 (3, September): 273–98.

———. 1990. "The Business Cycle Associated with Exchange Rate-Based Stabilizations." PRE Working Paper 513. World Bank, Country Economics Department, Washington, D.C. Processed.

Krugman, Paul, and Lance Taylor. 1978. "Contractionary Effects of Devaluation." *Journal of International Economics* 8 (3): 445–56.

Lizondo, Saul J., and Peter J. Montiel. 1989. "Contractionary Devaluation in Developing Countries: An Analytic Overview." *IMF Staff Papers* 36 (1): 182–227.

Modiano, Eduardo M. 1988. "The Cruzado First Attempt: The Brazilian Stabilization Program of February 1986." In Michael Bruno, Guido Di Tella, Rudiger Dornbusch, and Stanley Fischer, eds., *Inflation Stabilization*. Cambridge, Mass: MIT Press.

Obstfeld, Maurice. 1985. "The Capital In-Flows Problem Revisited: A Stylized Model of Southern Cone Disinflation." *Review of Economic Studies* 52 (4): 605–26.

Ortiz, Guillermo. 1991. "Mexico Beyond the Debt Crisis, Towards Stabilization and Growth and Price Stability." In Michael Bruno, Stanley Fischer, Elhanan Helpman, and Nissan Liviatan with Leora (Rubin) Meridor, eds., *Lessens of Economic Stabilization and Its Aftermath*. Cambridge, Mass: MIT Press.

Pazos, Felipe. 1972. *Chronic Inflation In Latin America*. New York: Praeger.

Ramos, Joseph. 1986. *Neoconservative Economics in the Southern Cone of Latin America, 1973–83*. Baltimore, Md.:Johns Hopkins University Press.

Rodriguez, Carlos A. 1982. "The Argentine Stabilization Plan of December 20th." *World Development* 10 (9): 801–11.

———. 1984. "La Estrategia de Estabilización con Tipo de Cambio Flexible y Política Monetaria Activa." *Cuadernos de Economia (Chile)* 21 (August): 103–21.

Sachs, Jeffrey D. 1986. "The Bolivian Hyperinflation and Stabilization." NBER Discussion Paper 2073. National Bureau of Economic Research, Cambridge, Mass. Processed.

Sargent, Thomas J. 1982. "The End of Four Big Inflations." In Robert E. Hall, ed., *Inflation: Causes and Effects*. University of Chicago Press.

———. 1986. "Stopping Moderate Inflations: The Methods of Poincare and Thatcher." *Rational Expectations and Inflation*. New York: Harper & Row.

Simonsen, Mario Henrique. 1974. "The Anti-Inflation Policy." In Mario Henrique Simonsen and Roberto de Oliveira Campos, eds., *The New Brazilian Economics*. Rio de Janeiro: Crown Editores Internacionais.

Taylor, John. 1979. "Staggered Wage Setting in a Macro Model." *American Economic Review Papers and Proceedings* 69 (2): 108–13.

Thorp, Rosemary. 1965. "Economic Policies in Argentina in the Post-War Years." *Bulletin of the Oxford University Institute of Statistics* 27 (2): 19–43.

Viana, Luis. 1988. "The Stabilization Plan of 1968." Banco Central del Uruguay. Processed.

Vicker, John. 1986. "Signalling in a Model of Monetary Policy with Incomplete Information." *Oxford Economic Papers* 38 (3): 443–55.

Wicker, Elmus. 1986. "Terminating Hyperinflation in the Dismembered Habsburg Monarchy." *American Economic Review* 76 (3): 350–64.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO 2: 307-330

A Framework for Evaluating the Impact of Pricing Policies for Cocoa and Coffee in Côte d'Ivoire

Pravin K. Trivedi and Takamasa Akiyama

This article presents an approach to evaluating pricing policies for perennial crops. A flexible computational model is developed, which incorporates important features of perennial crop production that are not captured by other (usually static) frameworks. This framework produces sensible and plausible scenarios for pricing cocoa and coffee in Côte d'Ivoire, as well as descriptions of revenue tradeoffs. Key issues arise from considering major changes in the rules used to set domestic producer prices. An unambiguously best policy is not determined, but several policies improve substantially on the present situation. Most of these alternatives indicate the desirability of lowering the tax on coffee relative to cocoa.

Many low-income developing countries depend heavily on tree crops, or perennials, for export and tax revenue, so that the health of the perennial sector and the overall state of the economy are closely connected. The most important perennials are coffee and cocoa, but rubber, tea, and oil palm are sometimes also significant. It is expected that most of these countries will continue to depend heavily on these crops for export tax revenue, and many will retain policies fixing both producer prices and exchange rates. Thus the choice of the appropriate tax rate and producer price, given government objectives, will remain of key importance. The research reported here evaluates alternative tax and price policies quantitatively, with specific reference to cocoa and coffee production in Côte d'Ivoire. Côte d'Ivoire was chosen for this analysis because it currently faces serious economic problems, whose resolution is closely entwined with its pricing policies for cocoa and coffee.

The objective of the study was to develop an operational approach for quantitatively evaluating pricing and tax policies for perennials over the short and

Pravin K. Trivedi is with the Department of Economics at Indiana University, Bloomington, Ind., and Takamasa Akiyama is with the International Economics Department at the World Bank. The authors acknowledge the interest, advice, and help of Ron Duncan at all stages of research. The current revised version has benefited greatly from the comments and suggestions for improvement received from two referees. Sanjiv Jaggia, a graduate student at Indiana University, provided valuable programming assistance for which the authors are grateful.

medium run (7 to 15 years). An "operational" approach is one that is computationally implementable and that quantifies the implications for production, exports, and tax revenues of different price rules under alternative assumptions about related variables such as the exchange rate, age-yield profiles, new plantings, supply elasticities, and minimum feasible producer prices. Because of the dynamics of perennial supply response, projections should cover a time horizon of 10 years or more. The revenue implications of a policy change are not fully revealed over a shorter period, so that focusing on short-term revenue flows will lead to myopic policies. Long-term projections, however, can be made only conditionally on projections of world demand and supply for the crop.

The framework developed here is an alternative to other approaches, including both ad hoc and partial evaluations concerned largely with the short-term impact and other model-based equilibrium approaches that do not project the time paths of important variables. This study is limited in scope; it does not consider dynamic price uncertainty in the face of credit constraints and attempts by governments to stabilize prices to overcome such problems, nor does it consider the policy credibility and reputation of the policymakers.

Section I provides a brief literature review of several alternative frameworks. Section II contains the essential institutional and factual background concerning pricing policies for cocoa and coffee in Côte d'Ivoire. Section III provides an overview of the structure of the model and discusses the measurement of the impact of policy changes on welfare. Section IV summarizes simulations of the model and gives the results of model sensitivity exercises. Section V discusses the choice between alternative policies, and Section VI concludes.

I. A Review of the Literature

Alternative pricing policies for perennial crops are evaluated in two stages. First, theoretical and computing frameworks must be developed to generate time paths for the relevant critical variables. Second, a criterion for choosing among alternative policies that differ in their impact on various sectors of the economy must be established. The second aspect is conceptually more difficult because it involves many of the classic normative issues of applied welfare analysis.

The first stage requires a careful integration of a priori assumptions and factual information. Broadly speaking, there are three approaches to developing an analytical framework: the static, general-equilibrium approach related to the optimal tax literature (Stern 1987); the static, partial-equilibrium approach rooted in the optimal tariff literature (Repetto 1972; Imran and Duncan 1988); and the dynamic, multimarket, partial-equilibrium approach of this study. Each of these approaches requires different types of inputs and provides different kinds of output.

The analytical framework used by Repetto and by Imran and Duncan is fairly straightforward, but it necessitates some strong assumptions, including that of a constant supply elasticity. This approach does not shed light on issues of income

distribution between cocoa and coffee producers. It is complemented by other analyses, such as that of Deaton and Benjamin (1988), who examine the desirability of adjusting the relative producer price of coffee and cocoa in Côte d'Ivoire by bringing them both in line with world prices. On average, the administered coffee price has been around 36 percent of the world price, and the cocoa price has been about 47 percent of the world price. Deaton and Benjamin consider the outcome of this price adjustment under several different assumptions about the degree of linkage between world and domestic producer prices and the degree of risk sharing between producers and the government. But they (correctly) remain neutral about the likely consequences in the absence of greater knowledge of producer response. They emphasize, as others have before them (see Newbery 1987), that supply response is crucial to the issue; however, their own framework and method is of a partial-equilibrium type.

In discussing the same set of issues for Côte d'Ivoire, Akiyama (1988) takes the approach that is closest to that of this study. The new-planting and replanting responses of cocoa and coffee producers under alternative tax rules are specified in detail, and potential production is calculated using the vintage matrix approach previously used in studies of perennials by Akiyama and Bowers (1984) and Akiyama and Trivedi (1987a, 1987b). The world price of the crop is projected using reduced-form versions of global commodity models, and the tax revenue consistent with both period-by-period price equilibration in the world market and the exchange rate policy being pursued is calculated. The model is simulated for each year of the relevant time horizon to generate the long-term consequences of a policy for producers' revenue, government revenue, and export revenue.

This computational framework, however, does not allow for the comparison of alternative policies or the evaluation of possible tradeoffs between producers' incomes and government tax revenue with different policies. Because Akiyama's approach does not use an algorithm to make relevant comparisons and does not calculate producer welfare, it needs to be extended and refined. Furthermore a better computational framework is needed to permit greater speed and flexibility in evaluating policy. Flexibility is especially important because developing-country data are frequently fragmentary and of poor quality so that they cannot support precise estimation of key structural parameters. Consequently, sensitivity and fragility analyses are essential.

II. Cocoa and Coffee Production in Côte d'Ivoire

Cocoa and coffee together have accounted for about 50 percent of total exports, 40 percent of agricultural gross national product, and more than one-half of the employment of Ivoirien farmers in recent years. Although the contribution of cocoa and coffee to the Ivoirien economy was quite significant during the 1970s, the weakening and subsequent collapse of world prices of the two crops after 1985 and the sharp appreciation of the CFA franc relative to the U.S.

dollar after 1985 dramatically changed the picture. Government revenues declined sharply, and, because of aggressive exchange rate policies pursued by Côte d'Ivoire's competitors, the country's international competitiveness weakened.

These changes have motivated a reexamination of current policies. The main alternative policy under consideration would reduce administered producer prices and relative producer prices of the two commodities so that prices would more accurately reflect marginal export revenues. The future world price outlook for coffee is brighter than for cocoa; therefore, such a price restructuring implies significant costs in terms of foregone producer and tax revenues.

The Caisse de Stabilization et de Soutien des Prix des Produits Agricoles (CSSPPA) controls the marketing of cocoa and coffee primarily through two policy instruments. First, at the beginning of each crop year it specifies the payment system (bareme), which details producer prices and domestic and external marketing costs along the entire marketing chain. Second, it uses an export price reference system to stabilize the price received by exporters. Exporters receive the world market price, but, if the world price exceeds the reference price, the exporter pays the CSSPPA the difference between the two. If the world price is lower than the reference price, the exporter is paid the difference between the two by the CSSPPA, which draws from its financial reserves.

The "Caisse" system has significant limitations. It is clear that the system will perform poorly when the CFA franc appreciates or when the world price declines for an extended period, because these circumstances necessitate prolonged revenue outflows from the CSSPPA. The resulting liquidity problems may force delays or suspension of payments to producers, and eventually the producer price may have to be reduced. The CSSPPA has run large deficits since the mid-1980s, and, for the 1989/90 crop year, producer prices for cocoa were reduced by 50 percent and for coffee by 40 to 50 percent.

Administered producer prices can create a domestic relative price structure that is out of alignment with domestic production costs if changes in these costs are not taken into account. This has occurred in Côte d'Ivoire; cocoa was priced too high relative to coffee, thus leading to large increases in cocoa production. Recent economic analyses of pricing policies have typically concluded that producer prices need to be realigned to eliminate the difference in favor of cocoa. Any such change will affect the distribution of income, so that an informed discussion of the merits of altering current relative prices must consider detailed information on the current income distribution among Ivoirien cocoa and coffee producers as well as the links between changes in relative prices and the distribution of income.

Deaton and Benjamin (1988) have examined the links between prices and farmers' incomes by using data from the 1985 Living Standards Measurement Study. Their analysis emphasizes several facts that have an important bearing on the interpretation and use of the simulation results presented below. First, cocoa and coffee farmers are not mutually exclusive groups. In fact coffee farmers

derive more net income from cocoa than from coffee. Second, cocoa farmers have average household incomes close to the national average, whereas those of coffee farmers are 13 percent below the average. Third, agricultural income constitutes about 65 percent of total income for these farmers, and cocoa and coffee account for about a third of agricultural income on average. Home-produced food is the most important source of income, accounting for 44 percent of net agricultural income.

III. Structure of the Model

The Ivoirien cocoa and coffee sectors are represented within the model by a detailed specification of output supply. Because both crops are cultivated in the same area, it is reasonable to postulate considerable output substitution between cocoa and coffee by farmers. Furthermore it is reasonable to postulate that substitution between annual food crops and perennials may be especially important. Because of the lack of data, however, this latter possibility has been ignored here (see Weaver 1989). Cocoa and coffee are viewed as substitutes from the producer's viewpoint, and the interaction between the two sectors is embedded in parametrically specified short-run supply equations and new-planting equations. It is assumed that the entire output of these sectors is exported. The interaction of the two sectors with the government is represented by a set of equations linking export prices with producer prices and export revenues with production and world prices as well as equations detailing the government's revenue and cost structure. The model does not include the interaction between the cocoa, coffee, and government sectors, on the one hand, and the rest of the Ivoirien economy, on the other. In other words the model can evaluate only the impact of macro policies, such as exchange rate adjustment, from a sectoral viewpoint.

The model used is nonlinear but essentially recursive and hence easy to solve. In broad terms the calculations proceed as follows. First, a regime for cocoa and coffee producer prices is specified together with assumptions about the exchange rate, age-yield profiles for both crops, and the time paths of Ivoirien production and world prices in the absence of production shocks in Côte d'Ivoire. An example of a policy regime is a cut of 30 percent in the producer prices of cocoa and coffee from base levels. It is also assumed that producers believe the announced price policy to be permanent and nonreactive and that they base their expectations and new-planting and production plans on the announced policy. That is, there are no issues of government credibility in the discussion.

Second, given producer prices, the model determines new plantings and the potential and actual production of cocoa and coffee. Third, given the cocoa and coffee supply curves, the production and profits of producers can be immediately obtained because, under the export price reference system, they do not depend upon yet-to-be-determined world prices. Next, together with world supply and demand, the policy-induced deviation of Ivoirien production from its

base production determines the deviation of the world price from its base. F specified exchange rate this then determines the Ivoirien export revenue a given the cost structure for marketing and transportation, the size of gove ment deficit or surplus in any given year. One simulation consists of solving t model, year by year, for a time horizon of up to 14 years (1987–2000) und specified initial conditions and a selected policy rule. To compare alternati policies, the results are aggregated for the entire horizon; to study the impact c any one policy, the results are presented as a time series. All ingredients for welfare evaluation are available at this stage.

The model is calibrated using historical data and parameter estimates that ar subject to considerable uncertainty because of the limited time series data used i their estimation. It should be emphasized that data constraints preclude the us of a fully econometric model. A not-inaccurate description of the model woul be that it is a numerical model that has been calibrated using econometri evidence and a priori restrictions. Most of the important restrictions have bee subjected to sensitivity analysis, which will be noted where appropriate.

The Supply Sector

The heart of the model is the specification of cocoa and coffee supply. Fo each crop, the block of supply equations is based on the vintage capital ap- proach, which uses data on new plantings, the age-yield profile, and death rate of trees by age group. A distinction is drawn between high-yielding and tradi- tional varieties. For the assumptions made in constructing vintage matrices, se Akiyama (1988).

A simple sketch of the supply sector is as follows. Let $Q^p(t)$ denote aggregat potential or expected average output and $Q(t)$ denote actual output, at time t, where $Q^p(t) \geq Q(t)$. Actual output is related to potential output via the short- run output supply equation

$$(1) \qquad Q(t) = g_1[Q^p(t)]g_2[p^d(t-i)], \qquad i = 0 \ldots n$$

where g_1 is a measure of the percentage of potential production that is actuall produced in any year, and the second term $g_2 p^d(t - i)$, represents the curren and lagged effects of own and substitute output prices, denoted by the vector p^d. Aggregate potential output is the sum of potential output from each of th different surviving age classes, assuming that the productivities of these cohort remain at their historic level, that is,

$$(2) \qquad Q^p(t) = \Sigma_v Q^p(t, v)$$

where v denotes the vintage or age class.

Given an age-yield profile, $\delta(t, v)$, $Q^p(t, v)$ depends upon the size of th surviving stock of vintage v trees, $K(t, v)$, through a production function relatio

$$(3) \qquad Q^p(t, v) = \delta(t, v)K(t, v)$$

where $K(t, v)$ is given by the accumulation equations

(4) $$K(t, v) = K(t - 1, v) - R(t, v)$$

(5) $$K(t, t) = N(t)$$

where $R(t, v)$ denotes removals, death, and losses from the stock exiting at $t - 1$ and $N(t)$ denotes new plantings. New planting of cocoa is a function of real producer cocoa and coffee prices, and the equation for new plantings of coffee is specified analogously:

(6) $$N(t) = N[p^d(t - 1)]$$

where $[p^d(t - 1)]$ is a proxy for the expected future real producer prices. (It would be desirable to restrict the form of the new-planting equation, for example, to impose constant returns to scale, but this would require data on planted areas by age class (Trivedi 1988), which are unavailable for Côte d'Ivoire.) The new-plantings specification implies static expectations, which is a strong assumption, but which has considerable appeal in the context of a study that explores the revenue and cost implications of particular fixed producer price regimes.

Calibration of the Model

Potential production in the model is determined by the stock and age composition of the trees. Figures for these do change, of course, as new planting occurs. Attempts at estimating these new planting equations using time series were summarized in Akiyama (1988). Although econometrically reliable estimates are difficult to obtain, there is evidence that aggregate new plantings respond to changes in real prices of the two commodities. Hence the equations used are based on the assumption that the two crops compete and that new planting in each sector responds to the relative price ratio. The best estimates for the own-price elasticity of new plantings of coffee are around 4.0 and for the cross-price elasticity are around −1.5. For cocoa the corresponding elasticities are 1.9 and −0.6, respectively.

On a priori grounds, however, the model uses variable, rather than fixed, new-planting elasticities. The rationale for this has been given elsewhere (Trivedi 1988), the essential idea being that the new-planting response to relative prices will be nonlinear in real prices. At a low relative price the elasticity will be small, and at a high relative price it will be high. The nonlinearity is calibrated to produce own- and cross-price elasticities similar to those given above at the real prices prevailing in 1986. When simulations incorporate a period of comparatively low real prices, however, planting elasticities are significantly reduced. Although it would be desirable to specify new-planting elasticities that decline as the supply of suitable land declines and as additional expansion uses agronomically inferior land, most of the policy options explored involve expan-

sion in only one of the two sectors, usually coffee. Given the assumption that the two crops compete, it seems consistent to assume that the decline in one sector releases land usable in the other so that expansion is not limited or halted by the exhaustion of suitable land.

The supply equations used for cocoa and coffee contain a multiplicative factor that depends upon feasible or potential production. The construction of vintage capital matrixes allows one to estimate the feasible production from each age cohort using a fixed age-yield profile. Aggregating these provides a measure of potential (expected average) production in any period. Assuming that potential and actual production differ by a scale factor, this scale factor may be obtained by regression of actual on potential production. The scale factors thus determined for cocoa and coffee are 1.02 and 0.99, respectively.

Time series regressions yield imprecise estimates of (Marshallian) own- and cross-price supply elasticities. Akiyama (1988) used point estimates of 0.22 for own-price elasticities of both coffee and cocoa, 0.08 for cocoa-coffee cross-price elasticity, and 0.04 for coffee-cocoa cross-price elasticity, these being close to values obtained using a short time series. Although the assumption of constant output supply elasticities is restrictive, the data needed for estimating a more flexible elasticity specification are not available. Instead an a priori belief is imposed that the supply response decreases as the actual price approaches the "shut-down" price, below which no production will take place. The 1982/83 average cost of production for coffee and cocoa for different plantation types (de Graaff 1986, table 9.8) is used as the minimum price below which production will not occur. The elasticities are specified as functions of prices such that at the real prices prevailing in 1986 the implied elasticities are close to the time series estimates given above, but at higher or lower prices their values change. These functions are given in table 1. The actual numerical values in those functions with such properties were obtained by a grid search of elasticities around the sample mean value estimated using a time series regression.

In the context of a static market-clearing model of demand and supply, the (absolute value of the) demand elasticity for Ivoirien production, denoted by E^d, is given by

$$E^d = [E^{w,d} + (1 - s)E^{w,s}]/s \qquad (7)$$

where $E^{w,d}$ and $E^{w,s}$ denote the rest-of-the-world demand and supply elasticities and s denotes the Ivoirien share of the world market. Let the domestic price, P^d, and the world price, P^w, be linked by $P^d = (1 - T)P^w$ where T is the export tax rate; then the relation between the (policy-induced) domestic price perturbation and the change in the world price is

$$dP^w/dP^d = -[1/(1 - T)]\{SE^s/[E^{w,d} + (1 - s)E^{w,s}]\} \qquad (8)$$

where E^s is the domestic elasticity of supply.

The world price of coffee is taken as exogenous, because the average Ivoirien share of world exports is only around 5 percent. The world price of cocoa,

Table 1. *Production and Planting Elasticity Equations*

$EP1CF = \exp(-BCF * MCF) * \exp(BCF * RFP)$
$EP2CF = 0.0012 * RFP$
$EP1CC = \exp(-BCC * MCC) * \exp(BCC * RCP)$
$EP2CC = 0.00058 * RCP$
$ETA1CF = 1.0 + 0.0434 * RFPMA/3$
$ETA2CF = 0.0236 * RFPMA/3$
$ETA1CC = 1.0 + 0.013 * RCPMA/3$
$ETA2CC = 0.00869 * RFPMA/3$

Notation:
MCC: Estimated minimum cost of production of cocoa/kg
MCF: Estimated minimum cost of production of coffee/kg
RFP: Real producer price of coffee
RCP: Real producer price of cocoa
RFPMA: Lagged 3-period moving sum of RFP
RCPMA: Lagged 3-period moving sum of RCP
EP1CF: Own-price production elasticity of coffee
EP2CF: Cross-price production elasticity of coffee
EP1CC: Own-price production elasticity of cocoa
EP2CC: Cross-price production elasticity of cocoa
ETA1CF: Own-price new-planting elasticity of coffee
ETA2CF: Cross-price new-planting elasticity of coffee
ETA1CC: Own-price new-planting elasticity of cocoa
ETA2CC: Cross-price new-planting elasticity of cocoa
$BCC = -1.514/(69 - MCC)$
$BCF = -1.514/(65 - MCF)$

however, cannot be treated as exogenous, because Côte d'Ivoire has close to 30 percent of the world share of cocoa exports. Ivoirien production and the world price are linked using a static market-clearing model of the world cocoa market. The rest-of-the-world supply elasticity is taken to be 0.30 and the demand elasticity −0.27, both close to those estimated in time series regressions. Furthermore, given a time series of the world cocoa price representing the base price, deviations from it can be computed using the above elasticities together with the assumption that a change in the world price arises when Ivoirien production deviates from its base level. The impact of an Ivoirien production shock on the world price is inversely related to the size of the shock relative to world production. The computation of the world price is done period by period; hence the generated perturbed solution incorporates the interdependence of the world price and policy-induced supply changes in Côte d'Ivoire.

Measures of Policy Impact and Welfare Evaluation

The key consideration in the evaluation of the impact of policy is the induced change in the distribution of income. The numerical model detailed above can be used to calculate the tradeoffs in terms of discounted net revenues that would

accrue to cocoa producers, coffee producers, and the government under alternative hypothetical pricing policies. An important qualification is that cocoa and coffee producers do not constitute mutually exclusive groups; therefore, they may be aggregated into a single "private sector," and the tradeoff between the government and the private sector can be considered.

Changes in producer prices directly affect production and profits in the cocoa and coffee sectors. At an aggregate sectoral level, the conventional measure of welfare change is the change in the producer surplus, or net profits of the producers. In the static production model, producer surplus is conventionally measured as the area bounded by the price and the supply function. Given the parametric specification of the supply function and the potential production for year t, the producer surplus, $\Pi(t)$, can be calculated as

$$\Pi(t) = \int Q(p^d)\, dp^d \tag{9}$$

where the range of integration is from the shut-down price to the government's current set price. This integral has an explicit, closed-form solution for the parameterization of the supply function used here and can be readily evaluated. A complication arises, however, from the dynamic adjustment process modeled here, which shifts the supply function intertemporally.

The presence of a multiplicative shift factor in the output supply equation means that as new planting occurs and potential output expands, the supply function shifts out, or the supply function is time-indexed. Aggregate producer surplus for the full time horizon is the discounted sum of the producer surplus in each year. The output supply function specified in equation 9 reflects only variable production costs and ignores the fixed costs of purchasing land and establishing a planted area, that is, the cost of capital inputs. This may be appropriate for a short-run analysis but will result in overestimation of producer surplus in the long run. To correct for this, Just, Hueth, and Schmitz (1981) suggest that the imputed value of the cost of preplanned capital inputs into current production can be subtracted from the static annual measure of producer surplus. This is roughly equivalent to subtracting the amortized value of capital costs from the producer surplus calculated in the conventional way and takes into account the costs of shifting resources into an expanding sector. Producer surplus was thus calculated as

$$\Pi(t) = \int Q(p^d)dp^d - Z(t) \tag{10}$$

where $Z(t)$ denotes annual amortized establishment cost of new plantings.

Most of the policies simulated here result in either an expansion of coffee production and a contraction of cocoa production or an increase in net government balance and a reduction in producer surplus. Choosing the optimum pricing policy, defined as that which maximizes aggregate social welfare, involves implicit or explicit weighing of the gains and losses of different groups. Explicit welfare weights will not be assigned here; instead the tradeoffs generated by alternative pricing policies will be presented. The choice of any point on the

tradeoff implies some social valuation of the relative gains and losses of all affected sectors. Deaton and Benjamin (1988) have concluded that the income distribution effects that resulted from shifts in the relative prices of cocoa and coffee were probably rather small. Assuming that this conclusion is valid, the key consideration in the choice of pricing policy is the social valuation of government tax revenue versus aggregate private revenue.

IV. Alternative Pricing Policies

Several simulations were carried out using the model, four of which are reported here. In the base case simulation it is assumed that the current pricing policies are continued. Then a policy of reducing both prices, along with a reduction in the relative price of cocoa, is simulated. Next the exchange rate is altered (the CFA franc is devalued), but absolute and relative producer prices remain unchanged. Finally, the effects of a system of nonfixed producer prices, which vary according to changes in the world price, are simulated. The projected time paths of relevant variables for these cases are illustrated in figures 1–6.

Base Case Projections and Simulation Results

The assumptions made to generate the base case projections are summarized in table 2, and the main results are as follows: annual cocoa production rises from about 620,000 tons in 1987 to 861,000 tons in 2000, while annual coffee production falls significantly, from 244,000 tons to 191,000 tons (figures 1 and 2). The cocoa sector's annual producer profit increases from CFAF146 bil-

Table 2. *Assumptions Underlying the Base Run*

Exchange rate (CFAF/$)	330.00
Producer price for cocoa	400.00
Producer price for coffee	377.00
World c.i.f. price	Endogenous
Domestic marketing and transportation (CFAF/kg)	111.20
Freight and insurance (CFAF/kg)	73.00
Export tax (CFAF/kg)	100.00
Cost to CSSPPA (CFAF/kg)	691.00
Cost to government	590.00
Cocoa: minimum production cost/kg (1987 CFAF)	45.725
Coffee: minimum production cost/kg (1987 CFAF)	67.14
World cocoa demand elasticity	−0.27
Rest-of-the-world cocoa supply elasticity	0.30
Annualized cocoa establishment cost/ha (thousands of 1987 CFAF)	16.80
Annualized coffee establishment cost/ha (thousands of 1987 CFAF)	22.40

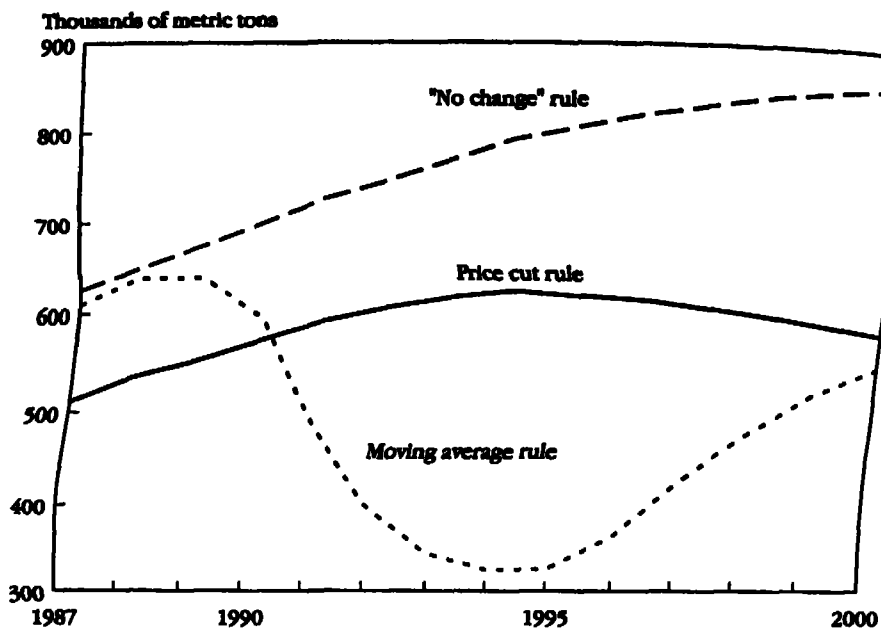
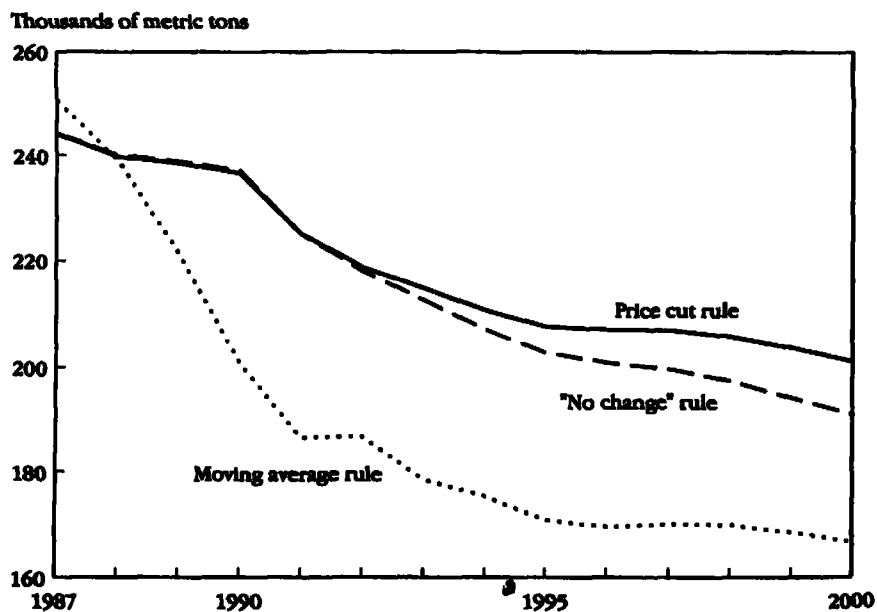
Figure 1. *Projected Cocoa Production under Three Price Rules*

Source: Authors' calculations.

Figure 2. *Projected Coffee Production under Three Price Rules*

Source: Authors' calculations.

lion[1] to CFAF203 billion, measured in 1987 prices. The cumulative surplus for cocoa is CFAF2,524 billion. For the coffee sector, annual producer profit falls from CFAF48 billion to CFAF34 billion, in 1987 prices. The cumulative surplus is CFAF537 billion. The net government balance from cocoa is in deficit in most years (figure 3), with the projected cumulative deficit for the full horizon being a massive CFAF1,957 billion. The annual deficit becomes smaller from 1992 onward, when the world cocoa price is projected to improve. For coffee the cumulative deficit is just CFAF26 billion (figure 4). Real cocoa export revenues are projected to increase from $743 million in 1990 to $1,340 million in 2000, while real coffee export revenues are expected to decline from $436 million to $406 million during the same period. Figures 5 and 6 show net producer revenue from cocoa and coffee. Quantitatively, these results indicate a considerably worse outlook than in Akiyama (1988), especially for government revenue (figures 3 and 4). The difference from Akiyama's results arises partly because lower average projected cocoa prices are used here. The continuation of current pricing policies implies massive deficits and, consequently, serious liquidity problems for the government.

Simulation of producer price changes at an unchanged exchange rate. The simulation reported here is for a policy of reducing the producer price of cocoa by 40 percent and of coffee by 10 percent. These reductions appear to be the minimum required, at an unchanged exchange rate, to produce modest government surpluses (of CFAF113 billion and CFAF90 billion for cocoa and coffee, respectively) for the full 14-year period and to lead to a diversification away from cocoa and into coffee and other crops.

Under this pricing policy, the sum of revenue for the government and profits for the cocoa and coffee producers for the full 14 years is CFAF2,292 billion compared with CFAF1,078 billion in the base run. Although this is an aggregate improvement, it is accompanied by massive changes in the relative sizes of the cocoa and coffee sectors, thereby precluding unambiguous statements about the welfare effects of such a policy. The cocoa sector's accumulated profit declines from CFAF2,524 billion in the base run to CFAF909 billion, while the coffee sector's accumulated profit increases from CFAF537 billion to CFAF1,179 billion. The net government balance from cocoa is not positive in every year, but for the period as a whole it is a surplus of CFAF113 billion, compared with an accumulated deficit of CFAF1,957 billion in the base run.

Cocoa production declines sharply with the change in relative prices, peaking at 623,000 tons in 1995 and declining to 585,850 tons in 2000. Relative to the base simulation, cocoa production is 38 percent lower by 2000. Côte d'Ivoire's share of the world cocoa market falls to 24 percent in the year 2000 compared with nearly 33 percent in the base run. Coffee production increases and by 2000 exceeds the base case by about 5 percent. As a consequence of the reduction in Ivoirien cocoa production, world prices throughout the period are significantly

1. A billion is 1,000 million.

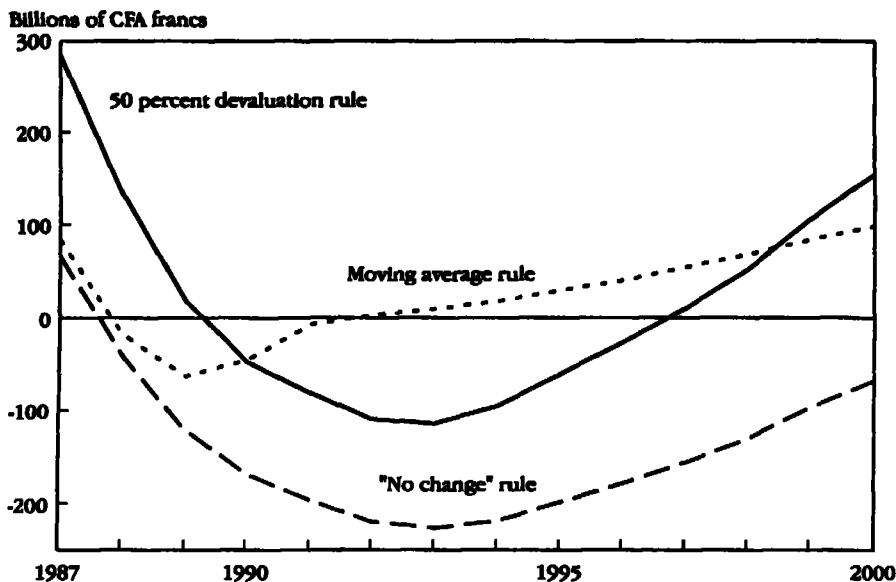
Figure 3. *Projected Net Government Balance from Cocoa under Three Price Rules*

Source: Authors' calculations.

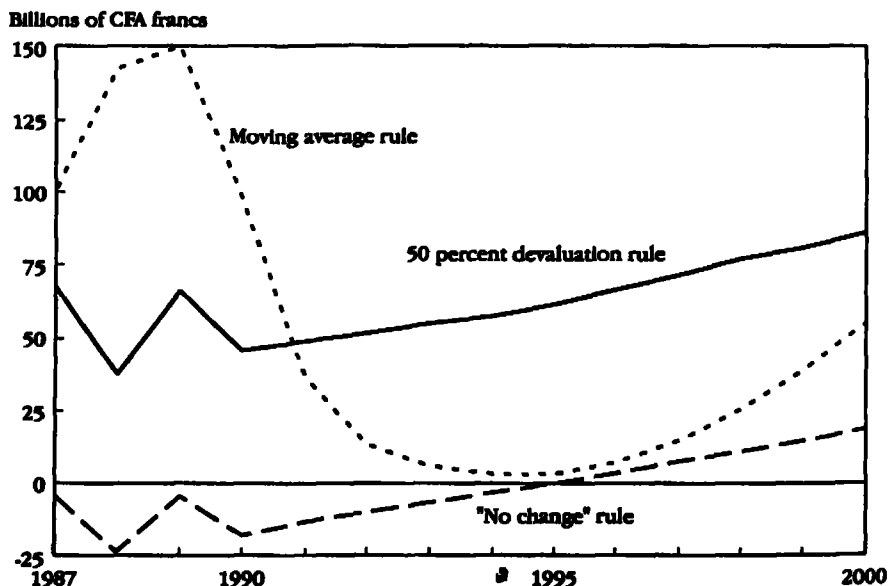
Figure 4. *Projected Net Government Balance from Coffee under Three Price Rules*

Source: Authors' calculations.

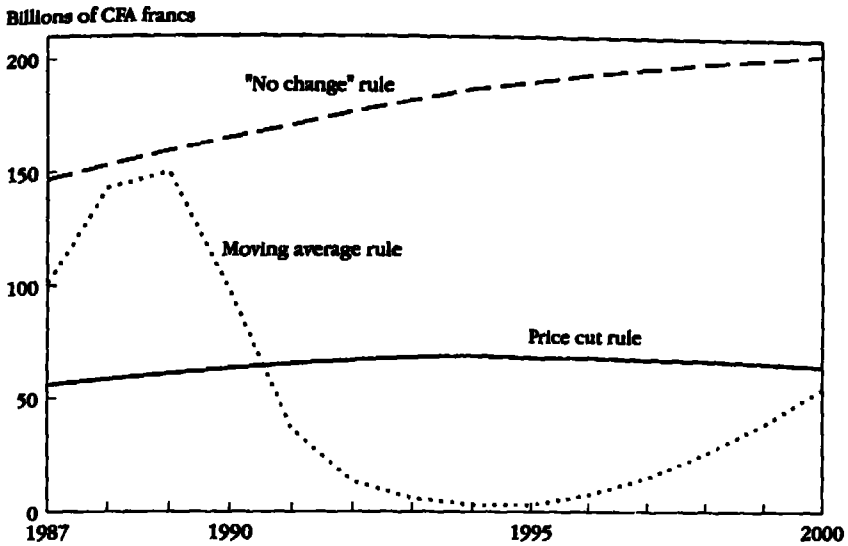
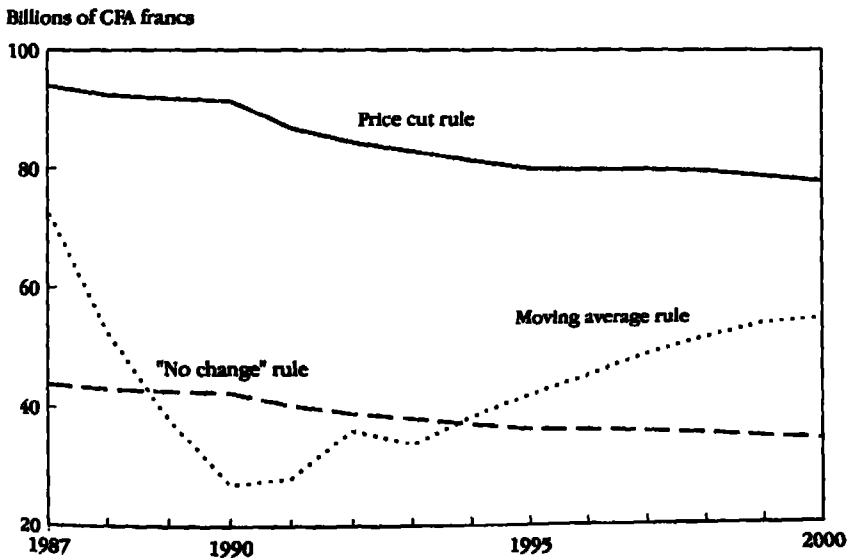
Figure 5. ***Projected Net Producer Revenue from Cocoa under Three Price Rules***

Source: Authors' calculations.

Figure 6. ***Projected Net Producer Revenue from Coffee under Three Price Rules***

Source: Authors' calculations.

higher than those in the base simulation, exceeding the base world price b percent in 2000.

Simulation of a 50 percent devaluation with no change in producer price. a 50 percent devaluation is adopted, there will be an immediate improvemer the net government balance. Because relative prices are unchanged, the size the cocoa and coffee sectors are unaffected. The cocoa surplus is nearly CFAF3 billion in 1987; it declines steadily (as the world price declines) to a deficit CFAF2.3 billion in 1992 and then improves steadily to a surplus of more th CFAF260 billion in 2000. Similarly, the coffee surplus also improves sharpl between 1990 and 2000 it rises steadily from about CFAF56 billion to abo CFAF94 billion. The cumulative surpluses over the simulation period a CFAF226 billion for cocoa and CFAF872 billion for coffee. Indeed the surpl would persist even if the coffee producer price were raised by 10 percent, 2 percent, or even 30 percent over the base case price. Therefore devaluation the CFA franc represents a powerful option for dealing with the problem persistent net government deficits.

Simulation of a moving average producer price system. To reduce produc price variability, some economists, for example, Mirrlees (1988), have suggest various forms of price adjustment in which changes in the producer price a some small fraction of the change in the expected world price; that is, produce are offered high, but not full, insurance. An example of such an adjustment ru is a moving average price in which the producer price equals a proportion of tl moving average of the expected world price.

The effects of instituting a three-year lagged moving average pricing rule we simulated, with producers receiving 70 percent of the average world price. strong assumption is made that the supply response and the rest of the mod will remain unchanged by such a switch in the pricing rule. Although the supp response may change, the absence of any historical experience makes it diffict to make concrete alternative assumptions. Without the price insulation prev ously provided by the government, the producer surplus and the governme deficit will be reduced when the market price falls below the price guaranteed the base run. This rule amounts to paying the producers in some years appro mately one-third of the price of cocoa in the base run. Relative to the base ru there is a sharp reduction in cocoa producer surplus and an increase in tl average annual coffee surplus. On average the government runs surpluses on i cocoa and coffee accounts.

Sensitivity Analysis

To examine the robustness of the conclusions of the simulations, sensitivi analysis was undertaken. Variations in the length of the policy horizon, the us of discounting when aggregating revenues over time, and variations of the as sumption about the shutdown price did not lead to significant changes in simul

tion results. The main focus of sensitivity analysis is on alternative settings of the key elasticities. However, the use of a simulation design in which the key elasticities are varied over a grid of values, one at a time, typically leads to a large volume of simulation output. To summarize this, a response surface was estimated. This involved estimating regression equations that described how the changes in certain key parameters generated changes in the variables of interest. The regression coefficients and their t-statistics indicated the size and statistical significance of the sensitivity. The key parameters in the present context were the sum of the absolute values of world supply and demand elasticities, the Ivoirien supply elasticities for cocoa and coffee, the planting elasticities for cocoa and coffee, and the percentage reductions in the producer prices of cocoa and coffee.

The sensitivity of key variables to parameter variation within and between two regimes—the fixed producer price (FPP) regime and the moving average price (MAP) regime—was investigated. For the FPP case, beginning with the base run setting, the sum of world supply and demand elasticities was varied from 0.10 to 0.60 in steps of 0.05, the cocoa price reduction was varied from 0 to 70 percent in steps of 10 percentage points, and the coffee price reduction was varied from 30 to 60 percent, also in steps of 10 percentage points. Thus 800 combinations were used. These changes also induced variation in Ivoirien planting and supply elasticities because these elasticities vary with the producer price, as described in tables 1 and 2. For the MAP case the procedure used was similar except that the percentage of the lagged three-year moving average world price that is paid to the cocoa and coffee producers was varied from 40 to 80 percent in steps of 10 percentage points—a procedure that, together with the 10 settings of the sum of world supply and demand elasticities, yielded 250 distinct parameter combinations. Post simulation, response surface regression equations were estimated with four dependent variables: aggregate net government balance from cocoa and from coffee and aggregate net producer profit from cocoa and from coffee. The regression results are shown in tables 3 and 4. A tight fit of the regression indicates that the included parameter settings provide a good explanation of the simulation variance of the variables of interest, and large t-statistics indicate that sufficient independent variation in parameter settings was allowed for in the simulation design.

In the fixed-price case (table 3) producers are sheltered from variations in the world price, so that only the net government balance from cocoa is sensitive to variations in the sum of world supply and demand elasticities. The higher this sum, the lower the tax yield is from a given reduction in producer price relative to the base run. Similarly, the higher the own-price supply elasticity, the lower the tax yield is for a given tax setting. Relative to the other elasticities, sensitivity to the planting elasticities was found to be quite small. The sensitivity of net government balance for cocoa and coffee to cross elasticities, while statistically significant in the case of cocoa (but not coffee) is smaller than for own-price elasticities. For the producer profit regressions, there is again significant sensitivity to the value of supply elasticities, with larger absolute values of these

Table 3. *Estimation Results for the Response Surface Regressions for Fixed Producer Price Regime*

Dependent variable	*Explanatory variable*					
	Percentage fall in cocoa price	*Percentage fall in coffee price*	*Sum of world supply and demand elasticities*	*Ivoirien cocoa supply elasticity*	*Ivoirien cocoa planting elasticity*	R^2
Net government balance from cocoa	1.02×10^4 (7.23)	-2.17×10^2 (9.83)	-1.16×10^3 (22.89)	-1.04×10^4 (45.86)	-1.89×10^3 (1.18)	0.978
Net government balance from coffee	−21.18 (0.026)	8.67×10^2 (83.51)	0.0 (0.0)	1.83 (0.0197)	−0.37 (0.0)	0.930
Net cocoa producer profit (in logs)	−12.90 (8.60)	1.51 (64.06)	0.0 (0.0)	−12.49 (61.59)	−19.42 (11.21)	0.980
Net coffee producer profit (in logs)	6.18 (1.08)	−4.72 (60.50)	0.0 (0.0)	−2.23 (4.15)	2.77 (0.43)	0.900

Note: Absolute values of *t*-statistics are in parentheses.
Source: Authors' calculations.

Table 4. *Estimation Results for the Response Surface Regressions for Moving Average Price Regime*

Dependent variable	*Explanatory variable*					
	World price of cocoa	*World price of coffee*	*Sum of world supply and demand elasticities*	*Ivoirien cocoa supply elasticity*	*Ivoirien coffee supply elasticity*	R^2
Net government balance from cocoa	-4.43×10^3 (72.07)	−94.09 (1.11)	2.40×10^2 (12.78)	-4.41×10^3 (55.70)	8.01 (0.08)	0.993
Net government balance from coffee	−37.48 (21.41)	-1.74×10^3 (495.00)	3.09 (3.48)	23.90 (9.45)	-1.55×10^3 (389.00)	0.999
Net cocoa producer profit (in logs)	0.47 (6.98)	−2.52 (17.68)	0.57 (18.34)	−5.82 (61.73)	1.20 (7.39)	0.997
Net cocoa producer profit (in logs)	−2.29 (12.32)	−4.13 (14.71)	0.51 (5.64)	−0.58 (2.40)	−13.23 (36.82)	0.994

Note: Absolute values of *t*-statistics are in parentheses.
Source: Authors' calculations.

elasticities leading to larger reductions in producer revenues when producer price is reduced.

The most obvious difference between the MAP case (table 4) and the FPP case is that in the MAP case, all government and producer revenues are sensitive to the sum of world supply and demand elasticities; larger values are associated with higher (not lower) revenues. Because the moving average rule generates a different time path of producer prices from the fixed price cases, however, caution must be exercised in comparing the regimes. Once the government no longer provides full-price insurance to producers, export tax revenues are less sensitive to the sum of world supply and demand elasticities, whereas producer revenues are considerably more sensitive. As in the fixed price case, government and producer revenues are again sensitive to the domestic supply elasticities.

The response surface estimates indicate that comparisons of producer and government revenues under alternative tax rates in the MAP regime are sensitive to the elasticity assumptions. Only government revenues are sensitive to elasticity assumptions in the FPP regime, but even that sensitivity is greater in the MAP case relative to the FPP case. This conclusion does not invalidate the exercise undertaken here but warns of the importance of the numerical assumptions in this article.

V. The Choice between Alternative Policies

In choosing a pricing policy, several tradeoffs must be considered. First, if the relative producer price is varied in Côte d'Ivoire, there will be short- and long-run changes in the producer revenues in the two sectors. The difference between short- and long-run changes reflects the latter's incorporation of the supply response. Second, there will be both a direct and an indirect impact on short- and long-run government revenues; the direct impact comes from changes in production and exports at the constant world price, and the indirect impact comes from a change in the world price resulting from the change in Côte d'Ivoire's share of world cocoa production. To illustrate the tradeoffs, two cases were considered. In the first the cocoa producer price was cut to 40 percent below the base price and kept at that lower level, while the coffee price change was varied between a cut of 60 percent and an increase of 30 percent. In the second case the coffee price was cut to 10 percent below the base price and kept at that lower level, while the cocoa price cut was varied between 0 and 70 percent. Figures 7 and 8 show the effect of each combination of price changes on sectoral revenues during the 14-year time horizon. From these figures one can easily infer the tradeoff between the revenues of the two producing sectors and the tradeoff between the government revenue, on the one hand, and total producer revenue, on the other. Because the net government balance schedule is steeper in figure 8 than in figure 7, we can infer that revenue is more readily affected by a given change in the price of cocoa rather than of coffee. Furthermore, because devaluation of the CFA franc will shift the net government balance

Figure 7. *Aggregate Sectoral Revenues with Cocoa Price Cut Held at 40 Percent*

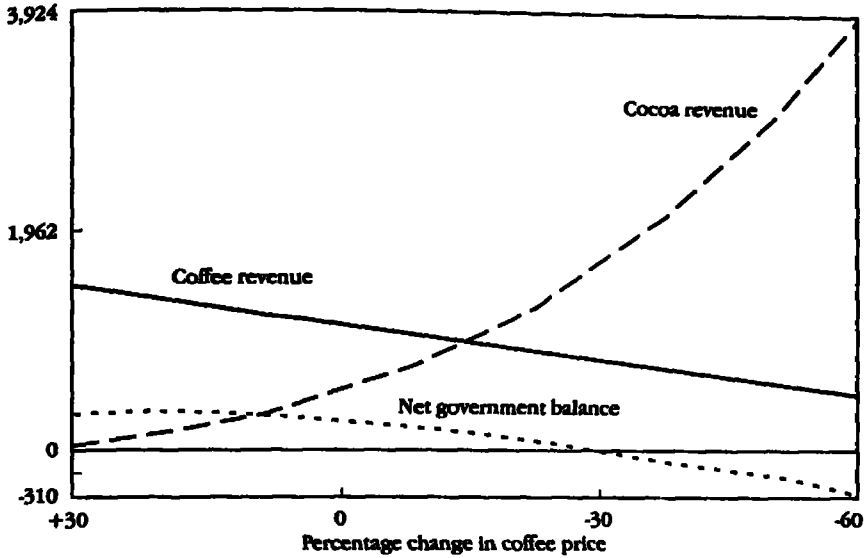
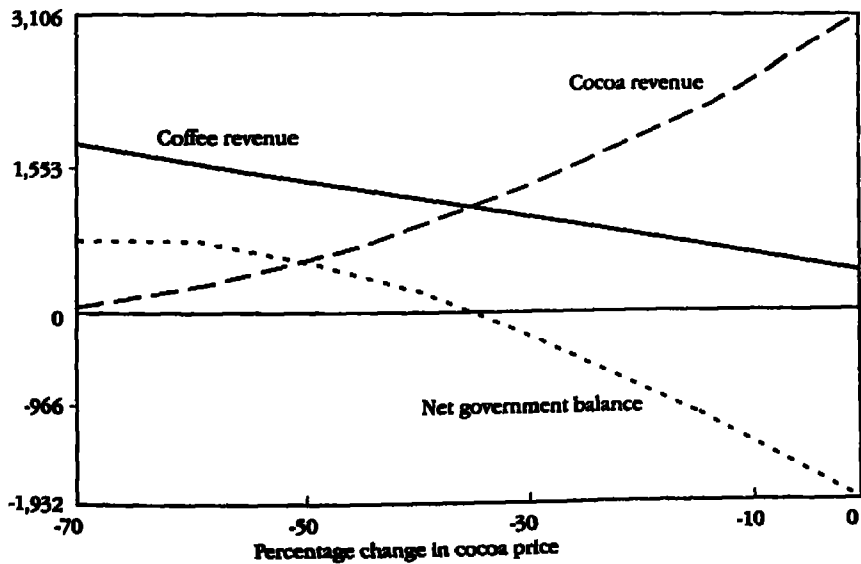
Billions of CFA francs

Source: Authors' calculations.

Figure 8. *Aggregate Sectoral Revenues with Coffee Price Cut Held at 10 Percent*

Billions of CFA francs

3,106
1,553
0
-966
-1,932
-70
-50
-30
-10
0
Cocoa revenue
Coffee revenue
Net government balance
Percentage change in cocoa price

Source: Authors' calculations.

schedule to the right, it will always improve the tradeoff between government and producer revenues.

In analyzing these tradeoffs, however, the possibilities for substituting alternative productive activities for cocoa and coffee production must be considered. There is a danger of exaggerating the impact of pricing policy on producers' welfare. Analogously, there is a danger of underestimating the impact on net government revenues as a consequence of producers shifting out of cocoa and coffee because of (say) reductions in prices. The empirical importance of this point depends upon the extent of substitute activities.

The analysis offered here may be combined with the optimal tax approach, in which the optimal tax rate is the reciprocal of the country's demand elasticity for a commodity, defined earlier as E^d. With Côte d'Ivoire's world cocoa market share at 0.3, the world elasticity of demand at 0.25, and the world elasticity of supply at 1.2, the optimal cocoa tax rate is estimated to be 27 percent. For coffee, if the market share is 0.04, the world elasticity of demand 0.25, and the world elasticity of supply 0.8, the optimal tax rate for coffee is 4 percent. (Because the period is 14 years, the supply elasticities are relatively large.) Thus this approach leads to substantially different tax rates for cocoa and coffee.

VI. Summary

This article has presented an approach to evaluating pricing policies for perennial crops. The main contribution is the development of a flexible computational model, which incorporates important features of perennial crop production that are not captured by other (usually static) frameworks. Applied to cocoa and coffee pricing in Côte d'Ivoire, the framework has produced sensible and plausible scenarios as well as useful descriptions of revenue tradeoffs. The study has highlighted many of the key issues that arise from considering major changes in the rules used to set domestic producer prices. It has not produced an unambiguously best policy but has identified several that improve substantially on the present situation, thus generally indicating the desirability of a lower tax on coffee than on cocoa.

The conclusions about Côte d'Ivoire policy are necessarily conditional on the assumptions about the future time paths of world cocoa and coffee prices and the choice of a 14-year policy horizon. The base scenario is characterized by massive deficits in the government cocoa account, rising cocoa production, and declining coffee production. If this scenario is to be avoided, clearly policy changes are needed. At an unchanged exchange rate, reducing the cocoa price about 40 percent and the coffee price about 10 percent would eliminate the government deficits on cocoa and coffee. To generate positive tax revenue, therefore, the price cuts would have to be larger. The devaluation of the CFA franc in conjunction with price cuts, or even on its own, would be a powerful way to reduce government deficits. This is, not surprisingly, a robust finding.

In October 1989 the Ivoirien government announced changes in the pricing of

cocoa and coffee, which, beginning with the 1989/90 growing season, would reduce the nominal cocoa producer price by 50 percent and the nominal coffee price by about 45 percent. A simulation of this policy shows that if it is maintained, the cocoa sector would continue to grow, but coffee production would decline even though the projected world price outlook for coffee is relatively more favorable than that for cocoa. Also, whereas the net government balance will be much improved, the deficit on cocoa will not be eliminated in every year, given the current projection of world cocoa price.

The scenario generated by the newly announced policy has some undesirable features, such as maintaining domestic relative prices or perhaps even distorting them further from their respective marginal export revenues. The consequences of this policy would be especially unfortunate if the projections of future cocoa and coffee prices do in fact materialize. The new policy may yield more tax revenue than would result from the adoption of the moving average price rule, with cocoa and coffee producers receiving 73 and 96 percent of the world price, respectively, but at the cost of a significant decline in coffee production. It is possible that aggregate producer and government revenues could be substantially lower. The simulation results based on the three-year moving average price rule show that such a policy effectively eliminates the large government deficits. However, since the projected outlook for cocoa prices is poor until the mid-1990s, the production of cocoa would decline very significantly under such a policy, even if the producers were paid a high proportion of the world price.

The system of setting producer prices in Côte d'Ivoire has parallels in many other countries and with many other crops (Varangis, Akiyama, and Thigpen 1989). Therefore the method and the results of this report have relevance extending beyond Côte d'Ivoire. Given a suitable data base, the present approach, suitably modified, can be applied to other countries with important perennial crops.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Akiyama, Takamasa. 1988. "Cocoa and Coffee Pricing Policies in Côte d'Ivoire." PRE Working Paper 64. World Bank, International Economics Department, Washington, D.C. Processed.

Akiyama, Takamasa, and Alan Bowers. 1984. "Supply Response of Cocoa in Major Producing Countries." Commodity Studies and Projections Division Working Paper 1984-3. World Bank, International Economics Department, Washington, D.C. Processed.

Akiyama, Takamasa, and Pravin K. Trivedi. 1987a. *A New Global Tea Model: Specification Estimation and Simulation*. World Bank Staff Commodity Working Paper 17. Washington, D.C.

———. 1987b. "Vintage Production Approach to Perennial Crop Supply." *Journal of Econometrics* 36 (1/2): 133–62.

Deaton, Angus, and Dwayne Benjamin. 1988. *Living Standards Survey and Price Policy Reform: A Study of Cocoa and Coffee Production in the Côte d'Ivoire.* LSMS Working Paper 44. Washington, D.C.: World Bank.

de Graaff, Jan. 1986. *The Economics of Coffee.* Wageningen: Pudoc.

Imran, Mudassar, and Ronald C. Duncan. 1988. "Optimal Export Taxes for Perennial Crop Exporters." PRE Working Paper 10. World Bank, International Economics Department, Washington D.C. Processed.

Just, Richard E., Darrell L. Hueth, and Andrew Schmitz. 1982. *Applied Welfare Economics and Public Policy.* Englewood Cliffs, N.J.: Prentice-Hall.

Mirrlees, James A. 1988. "Optimal Commodity Price Intervention." Oxford University, Department of Economics. Processed.

Newbery, David M. 1987. "Agricultural Taxation: The Main Issues." In Newbery, David M., and Nicholas Stern, eds., *The Theory of Taxation for Developing Countries.* New York: Oxford University Press.

Repetto, Robert. 1972. "Optimal Export Taxes in the Short and Long Run, and an Application to Pakistan's Jute Export Policy." *The Quarterly Journal of Economics* 86 (3): 396–406.

Stern, Nicholas. 1987. "The Theory of Optimal Commodity and Income Taxation: An Introduction." In Newbery, David, and Nicholas Stern, eds., *The Theory of Taxation for Developing Countries.* New York: Oxford University Press.

Trivedi, Pravin K. 1988. "A Model of Cocoa Replanting and New Planting in Bahia, Brazil, 1966–85." PRE Working Paper 92. World Bank, International Economics Department, Washington, D.C. Processed.

Varangis, Panyos, Takamasa Akiyama, and Elton Thigpen. 1989. "Recent Developments in Marketing and Pricing Systems for Agricultural Export Commodities in Sub-Saharan Africa." PRE Working Paper 431. World Bank, International Economics Department, Washington, D.C. Processed.

Weaver, Robert D. 1989. "An Integrated Model of Perennial and Annual Crop Production for Sub-Saharan Countries." PRE Working Paper 175. World Bank, International Economics Department, Washington, D.C.

THE WORLD BANK ECONOMIC REVIEW, VOL 6, NO. 2: 331-351

The Short- and Long-Run Effects of Fiscal Policy

Edward F. Buffie

This article develops a dynamic, dual-economy general equilibrium model that can be adapted to analyze the short- and long-run effects of a variety of fiscal policies. The model provides a complete description of how the private capital stock, underemployment, and real wages evolve during the adjustment process. The main policy message conveyed by the results is that the method by which the fiscal deficit is lowered is important. There is a strong presumption that higher prices for publicly produced intermediate inputs and cutbacks in government expenditure to support social infrastructure will reduce private investment, real wages in both the formal and informal sectors, and the share of the labor force employed in the high-wage manufacturing sector. By contrast, layoffs in the final goods and services sectors can potentially improve the external balance without sacrificing output and employment growth.

A central dilemma facing policymakers in developing countries today is how to revive economic growth while maintaining debt service. After the debt crisis in 1981–82, IMF-type stabilization programs were widely adopted. Austerity measures along with high interest rates and recession in the countries of the Organisation for Economic Co-operation and Development led to sharp reductions in real output in most debtor nations. The real side repercussions of macroeconomic austerity have cut deeper and been more damaging than was anticipated by many observers (Sachs 1989).

Two striking aspects of the adjustment after 1982 in many developing countries have been the collapse of investment (both public and private) and the sharp increase in underemployment. Table 1 provides data on real per capita growth rates of gross domestic product (GDP) and investment in 15 major debtor nations and in developing countries not burdened by debt service problems. The investment rate in the major debtor nations fell 32 percent between 1981 and 1984. It has increased slightly since 1984 but remains some 10 points below the average investment rate in countries not experiencing debt-servicing difficulties.

Low rates of growth and investment have been accompanied by rising underemployment. Public sector layoffs and stagnation in the industrial sector have

Edward F. Buffie is with the Department of Economics at Indiana University. He is indebted to Michael Walton, two anonymous referees, and the participants in two conferences on Labor Markets in an Era of Adjustment (at the University of Warwick) for helpful comments on an earlier draft.

Table 1. *Per Capita Growth and Investment, 1980–89*

	Per capita real GDP growth (percent)		*Investment (percentage of GDP)*	
Year	*Major debtors*	*Developing countries without debt servicing problems*	*Major debtors*	*Developing countries without debt servicing problems*
1980	3.1	3.5	24.2[a]	—
1981	−2.6	3.2	23.5	27.8
1982	−2.5	2.7	21.6	27.0
1983	−4.7	4.6	17.1	27.1
1984	—	5.4	16.0	27.0
1985	1.6	4.3	16.8	28.6
1986	1.9	3.8	17.2	28.3
1987	0.1	4.3	17.4	27.7
1988	−1.1	6.7	17.0	27.9
1989[b]	−1.4	3.2	17.3	27.5

— Not available.
a. Average figure for 1970–82.
b. Preliminary figures.
Source: International Monetary Fund (various years).

greatly slowed or brought to a halt employment growth in the principal high-wage sectors of the economy. It is particularly noteworthy that higher underemployment has often occurred in conjunction with large decreases in real wages. Tables 2 and 3 show that in several Latin American countries industrial employment declined during the 1980s despite real wage cuts of 10–30 percent.

Although the adjustment process has proven to be quite lengthy, there is little formal analysis in the existing literature of the long-run repercussions of stabilization policy. Few development macromodels afford a perspective that stretches beyond the short run (Arida and Taylor 1989). Public finance theorists have analyzed the interdependence of public and private investment but not in a context that sheds light on the adjustment problems facing developing countries. Most of this literature seeks only to determine the implications for the social discount rate of distortionary taxes and different types of government budget constraints (Boadway 1978; Pestieau 1974, 1975; Marchand, Pestieau, and Weymark 1982; Yoshida 1986). Moreover the orientation in the literature is largely static. Arrow and Kurz (1970) and Boadway (1978) are exceptions, but their dynamic analysis is based on dubious, ad hoc specifications for private investment. None of the literature allows for labor market distortions of the type seen in developing countries or develops an explicit dynamic analysis of capital accumulation consistent with optimizing behavior in the private sector.

This article develops a dynamic, dual-economy general equilibrium model that can be adapted to analyze the short- and long-run effects of a variety of fiscal policies. The dynamics are grounded in optimizing behavior and provide a complete description of how the stock of private capital, sectoral employment, and real wages evolves during the adjustment process. The focus is on fiscal policy because fiscal adjustment has been a prominent part of many recent

stabilization programs, and there appears to be a strong, direct link between labor market developments and certain types of fiscal policies.

The main policy message conveyed by the results is that the success of the adjustment program depends in large part on the method by which the fiscal deficit is lowered. Certain fiscal measures commonly adopted in highly indebted countries produce contractionary supply-side effects. There is a strong presumption that higher prices for publicly produced intermediate inputs and cutbacks in government expenditure to support social infrastructure will reduce employment in the high-wage manufacturing sector and lower the equilibrium capital stock. Labor bears the brunt of the difficult adjustment as lower employment in manufacturing is accompanied by (possibly large) real wage cuts in both the formal and informal sectors.

Table 2. *Industrial Employment in Highly Indebted Latin American Countries, 1981–88*
(1980 = 100)

Country	*1981*	*1982*	*1983*	*1984*	*1985*	*1986*	*1987*	*1988*
Argentina	87.4	82.8	85.5	88.0	84.7	81.3	81.0	82.1
Brazil[a]	90.7	85.3	79.0	78.1	79.1	86.9	89.7	88.2
Colombia	95.4	90.4	84.4	83.4	81.6	81.3	83.8	86.1
Chile	98.5	71.3	74.3	81.8	86.6	93.1	100.3	111.2
Costa Rica	94.6	92.7	96.2	99.3	99.8	100.4	97.7	104.4
Mexico	105.6	103.0	93.1	92.2	94.3	90.5	87.4	n.a.
Peru[b]	101.1	99.7	94.3	84.4	83.4	88.4	95.9	92.4
Venezuela	100.0	102.0	100.3	98.9	99.9	102.6	110.7	119.3

a. Industrial employment in the nine major metropolitan areas.
b. Manufacturing employment in the Lima metropolitan area.
Source: United Nations (1988).

Table 3. *Average Real Wages in Highly Indebted Latin American Countries, 1981–88*
(1980 = 100)

Country	*1981*	*1982*	*1983*	*1984*	*1985*	*1986*	*1987*	*1988*[a]
Argentina	89.4	80.1	100.5	127.1	107.8	109.5	103.0	97.3
Brazil[b]	108.5	121.6	112.7	105.1	112.7	121.8	102.4	107.1
Colombia	101.3	104.7	110.1	118.1	114.6	120.1	119.2	117.7
Chile	108.9	108.6	97.1	97.2	93.5	95.1	94.7	101.0
Costa Rica	88.3	70.8	78.5	84.7	92.2	97.8	89.2	87.5
Mexico	103.6	104.4	80.7	75.4	76.6	72.3	72.8	n.a.
Peru	101.8	110.2	93.4	87.2	77.6	97.5	101.3	77.4
Uruguay	107.5	107.1	84.9	77.1	88.1	94.0	98.5	99.7

Note: Figures are usually for workers employed in manufacturing or industry. For more detailed descriptions of the individual wage series, see the notes to table 11 in United Nations 1988.
a. Preliminary figures.
b. Average real wage in basic industry in Rio de Janeiro.
Source: United Nations (1988).

By contrast layoffs in the final goods and services sectors can potentially improve the external balance without sacrificing output and employment growth. The layoffs stimulate investment by returning real resources to the private sector. Private capital accumulation eventually creates enough new employment in the high-wage manufacturing sector to fully compensate for the loss of public sector jobs; in the long run, real output is higher and formal sector employment and real wages are unchanged.

The article is organized into seven sections. Sections I to III develop the basic model and analyze the impact of public sector price hikes and employment cuts. Section IV investigates the repercussions of reducing infrastructure investment. Section V contrasts the adjustment processes associated with the different fiscal measures and discusses the extent to which capital decumulation and lower employment in the high-wage sectors increase the cost of adjustment. Section VI examines how the results change when endogenously varying tax revenues make the fiscal and real adjustment mechanisms interdependent. Section VII expands on the broad policy implications of the analysis.

I. A Pure Supply-Side Model

The models in this and subsequent sections highlight the supply-side effects of fiscal austerity. To abstract from demand-side complications, I assume the economy is small and completely open. Two traded goods are produced: an agricultural export good and a manufactured good. The price of each good is fixed at unity. Production in the manufacturing sector requires labor, capital, and an intermediate input (such as gas or electricity) purchased from the public sector. The agricultural good is produced by just labor and land. Introducing capital and intermediates as factors in the agricultural sector does not substantively alter the results, provided the manufacturing sector is relatively capital- and intermediates-intensive.

Numerous empirical studies conclude that sectoral wage differentials in developing countries are far too large to be explained by the payment of compensating differentials (Gregory 1975; Merrick 1976; Squire 1981, chapters 7 and 8; Mazumdar 1976, 1989a, 1989b; House 1984; Portes, Blitzer, and Curtis 1986; Gindling 1989). The labor market is highly dualistic, with wages in the modern formal sectors sometimes being more than double those paid elsewhere in the economy. In keeping with the findings of these studies, the agricultural sector in the model is equated with the low-wage informal sector; government and manufacturing comprise the high-wage formal sector. A genuine labor market distortion thus exists because there is too little employment in private manufacturing relative to agriculture.

Although the sectoral wage gap generates underemployment, there is no open unemployment. All those unable to obtain work in the government or manufacturing sectors are employed in the agricultural sector, where the wage adjusts to clear the market. The labor market is represented by equation 1:

$$L^m + L^x + L^g = L \tag{1}$$

where L^m, L^x, and L^g denote employment in manufacturing, agriculture, and the public sector, respectively, and L denotes total labor supply. Total labor demand consists of private sector labor demand plus public sector employment. The total supply of labor is constant.

The most troublesome issue in modeling the labor market concerns the appropriate way to make the manufacturing sector wage endogenous. In theory it may be set in implicit contracts to provide insurance to workers by unions, by sociopolitical norms embodied in minimum wage laws, or by efficiency wage considerations. Unfortunately, empirical work on the wage-setting process in developing countries is scarce and does not single out one theory as clearly superior. Furthermore, although the aforementioned theories may explain wage rigidity in manufacturing, the only restriction they place on how the wage responds to various shocks is that, other things being equal, the manufacturing sector wage should be positively related to the agricultural sector wage. As neither theory nor empirical studies offer much guidance, I choose a particularly simple specification

$$\hat{w}^m = b\hat{w}^x, \qquad 0 < b \leq 1 \tag{2}$$

where w^m and w^x denote wages in manufacturing and agriculture, respectively, a circumflex indicates a percentage change in a variable, and b is constant (and positive to ensure the existence of a steady state). This specification is consistent with the Solow condition (when $b = 1$) in efficiency wage models and with certain variants of the optimizing union model. The parameter b plays a crucial role in the adjustment process because it determines the degree of real wage rigidity in manufacturing. If $b = 1$, the labor market is distorted by a sectoral wage gap, but both agriculture and manufacturing are flex-wage sectors. However, when b is small the real wage in the formal sector is largely impervious to economywide employment conditions. More of the burden of adjustment to contractionary policies is then borne by wage cuts in the informal sector and increases in underemployment (that is, greater layoffs in the manufacturing sector).

Firms are perfectly competitive and operate with technologies characterized by constant returns to scale. The zero profit condition is therefore satisfied in each sector

$$1 = C^m(w^m, r, P) \tag{3}$$

$$1 = C^x(w^x, v) \tag{4}$$

where C^m and C^x denote the unit cost function in the manufacturing and agricultural sector, respectively; r and v are the capital and land rentals; and P is the price of the intermediate input purchased from the public sector.

For simplicity I assume technology in each sector can be represented by a

(non-nested) constant elasticity of substitution (CES) production function. Private sector demands for manufacturing and agricultural labor and for the intermediate input (denoted by Z) are then (in percentage changes)

$$\hat{L}^m = -\sigma^m \frac{1-\theta_Z}{\theta_K}\hat{w}^m - \sigma^m \frac{\theta_Z}{\theta_K}\hat{P} + \hat{K} \tag{5}$$

$$\hat{L}^x = -\frac{\sigma^x}{\theta_T}\hat{w}^x \tag{6}$$

$$\hat{Z} = -\sigma^m \frac{\theta_L}{\theta_K}\hat{w}^m - \sigma^m \frac{1-\theta_L}{\theta_K}\hat{P} + \hat{K} \tag{7}$$

where σ^m and σ^x are elasticity of substitution in the manufacturing and agricultural sectors, respectively; θ_L, θ_Z, and θ_K are, respectively, the cost shares of labor, the intermediate input, and capital in the manufacturing sector; θ_T is the cost share of land in the agricultural sector; and K denotes capital. The only characteristic of CES technology that is important for the results that follow is gross complementarity of factors (that is, an increase in the price of the intermediate input lowers demand for labor and capital in manufacturing). This is not a particularly strong restriction to place on technology. According to production theory, factors are normally gross complements (Rader 1968). Empirical studies also find, with rare exceptions, that complementarity holds.

Capital accumulation is governed by factor returns and the intertemporal preferences of a representative, infinitely-lived family firm. The firm is endowed with perfect foresight and chooses investment to maximize an additively separable utility function

$$\max_{\{E,I\}} \int_0^\infty V(E)e^{-\rho t}dt \tag{8}$$

subject to

$$E + I = R(K, L^m, L^g, P) + w^g L^g - T \tag{9}$$

$$\dot{K} = I - \delta K, \tag{10}$$

where E is aggregate consumption expenditure, I is investment, T is a lump-sum tax, ρ is the pure rate of time preference, w^g is the public sector wage, δ is the depreciation rate, and an overdot signifies a time derivative. Current utility is represented by an increasing, strictly concave indirect utility function $V(\cdot)$.

Equation 9, the budget constraint, states that consumption and investment spending must equal disposable income. (Neither the private agent nor the government has access to foreign credit.) On the right side, private sector value added is measured by the value added function $R(\cdot)$, in which the fixed total supplies of labor and land are suppressed. The value added function has the

usual properties that an increase in the capital stock raises real output by an amount equal to the real capital rental and an increase in the real price of the intermediate input lowers output by an amount equal to the initial demand for intermediates. Also, since employment increases in other sectors of the economy come at the expense of employment in agriculture, higher public sector employment lowers private value added by an amount equal to the agricultural wage, while the marginal gain from expanding employment in manufacturing is measured by the existing sectoral wage gap $(w^m - w^x)$.

The government must respect the budget constraint

$$w^g L^g + D = PZ + T \tag{11}$$

where D denotes debt service. Total public sector expenditure is the sum of the wage bill and debt service (net of new capital flows). The profile of debt service is determined by negotiations with foreign creditors and is treated as strictly exogenous. The public sector wage is constant, and, of the L^g workers hired by the government, L_1 are employed in producing the intermediate input $[Z = Z(L_1)]$. But although L_1 is endogenous, the extent to which public sector employment varies with private sector demand for the intermediate input is a policy variable. When the demand for intermediates contracts, labor needed by the parastatal sector falls by $dL_1 = (Z/Z')\hat{Z}$, whereas the change in total public sector employment is

$$dL^g = \beta(Z/Z')\hat{Z}, \qquad 0 \leq \beta \leq 1. \tag{12}$$

β defines the government's layoff policy. When $\beta < 1$, redundant labor is kept on the payroll or transferred to other activities where it produces "government services." In either case the short- and long-run results and the qualitative nature of the dynamics are unchanged.

Government revenue derives from two sources: a lump-sum tax (T) and sales of the intermediate input (PZ). The unrealistic assumption of a lump-sum tax is made at this point to simplify the analysis. The impact on the budget of variations in private sector demand for the intermediate input is offset by adjustments in the lump sum tax so that higher debt service can be dealt with by a one-time adjustment in the price of the intermediate.

Equations 1 to 12 form the complete model. Since private sector saving and investment are equal, the trade balance is $PZ + T - w^g L^g = D$; thus the overall balance of payments equals zero. In what follows, debt service increases from an initial value of zero and a fiscal instrument is adjusted to extract the required trade surplus.

II. The Short- and Long-Run Impacts of Public Sector Price Increases on Output, Employment, and Real Wages

The manipulations involved in solving a perfect foresight general equilibrium model are straightforward but also lengthy and tedious. To keep the main ideas

clearly within view, the exposition here is mostly verbal and graphical. Solution procedures for the short- and long-run outcomes and proofs of saddle point stability may be found in a more technical version of the article that is available from the author upon request.

The Short-Run Impact

An increase in the price of the intermediate input lowers labor demand in private manufacturing and in the parastatal sector at existing wages. Aggregate high-wage employment thus contracts, forcing the agricultural wage to decrease.

Labor demand in the manufacturing sector is subject to two conflicting effects. While the decrease in the agricultural wage triggers a fall in the manufacturing sector wage, the higher price of intermediates shifts the labor demand schedule to the left. Employment rises or falls depending on whether

$$\sigma^m \gtrless \sigma^x \frac{L^x(1-s)\theta_L}{L^M\theta_T b\beta} \tag{13}$$

where $s \equiv (w^m - PZ')/w^m$ denotes the percentage gap between the marginal product of labor in private manufacturing and in the parastatal sector. If $b\beta$ is small either because the government maintains the level of public sector employment ($\beta = 0$) or because the manufacturing sector wage responds weakly to changes in the agricultural wage (b is small), the adverse productivity effect dominates, and manufacturing sector employment declines. More generally, employment in both high-wage sectors is likely to contract unless technology is far more flexible in manufacturing than in agriculture. Since the share of the labor force employed in private manufacturing is small (the ratio of agricultural to manufacturing sector employment generally lies between two and seven), the term multiplying the elasticity of substitution in the agricultural sector (σ^x) in expression 13 will usually be quite large.[1] Even when the government takes a tough line on layoffs ($\beta = 1$), real wages are equally flexible in the formal and informal sectors ($b = 1$), and the productivity gap between labor in the public and private manufacturing sectors is 50 percent ($s = 0.5$), the elasticity of substitution in the manufacturing sector has to be substantially larger than the elasticity of substitution in the agricultural sector in order for employment in the manufacturing sector to increase.

Dynamics and the Long-Run Impact

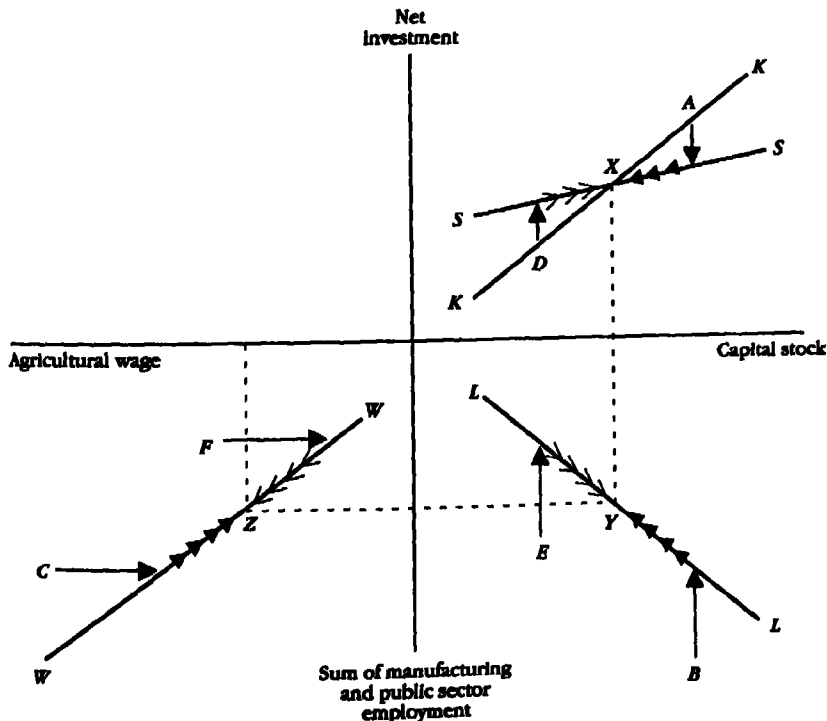
The adjustment process stretches beyond the short run because fiscal austerity affects the incentive to accumulate capital. As investment gradually alters the

1. In 1980 the employment share of the industrial sector (a rough proxy for the employment share of the formal sector) was 13 percent in low-income developing countries, 23 percent in middle-income developing countries, and 31 percent in upper middle-income developing countries. In our dual economy model, the corresponding values for L^x/L^m are 6.7, 3.4, and 2.3.

capital stock, the temporary equilibrium is displaced and further changes occur in real output, sectoral labor demands, and real wages.

The important qualitative features of the adjustment process are depicted in figure 1. The steady state is a saddle point with a unique convergent path to equilibrium. In the first quadrant the positively sloped *KK* schedule shows the set of points for which net investment is zero. Above *KK* net investment is positive and the capital stock is increasing; below the schedule, the capital stock is falling. The saddle path *SS* may be positively or negatively sloped. Regardless of the slope of *SS*, the capital stock approaches its steady-state level monotonically.

The *WW* and *LL* schedules in the third and fourth quadrants complete the description of the equilibrium path. These schedules track the paths of the agricultural wage and high-wage sector employment as the economy traverses the saddle path *SS*. Both schedules are positively sloped because an increase in the capital stock bids up the market clearing value of the agricultural wage by raising labor demand in the high-wage sectors.

Figure 1. *Major Qualitative Features of the Adjustment Process*

The dynamics of the adjustment process depend entirely on how the polic package affects the steady-state capital stock. Across steady states

$$\hat{K}/\hat{P} = \left[\sigma^m(s + \beta - 1) - \sigma^x \frac{L^x(1 - s)}{L^m\theta_T b}\right]\theta_Z/N \quad (14)$$

where $N \equiv \theta_L(1 - s) + \beta\theta_Z$.

The impact on the equilibrium capital stock depends on whether the increas in the price of the intermediate and the induced decrease in the manufacturin sector wage combine to raise or lower the profitability of investment. Althoug initial conditions (that is, L^x/L^m) and the nature of technology influence th outcome, equation 14 yields several well-defined results. First, for small b th potentially positive term involving the elasticity of substitution in the manufac turing sector (σ^m) is dominated by the negative term on the right side of equatio 14. Thus K declines when there is a high degree of real wage rigidity in th manufacturing sector because the manufacturing sector wage does not adjus enough to preserve profitability. Second, the capital stock always decrease when the parastatal sector "properly" belongs to the high-wage sector ($s = 0$) Third, capital decumulation occurs if the government resists making layoffs More precisely, the smaller the productivity gap between manufacturing an parastatal labor, the tougher must be the government layoff policy. There is n hope whatsoever of stimulating capital accumulation unless $\beta > 1 - s$.

In the most general case there is a strong presumption that the capital stoc will fall. Even when the productivity gap (s) is quite large, the governmen adopts a tough layoff policy, and the real wage in manufacturing is highl flexible, capital decumulation is to be expected. For the example considere earlier in which the government lays off workers strictly as dictated by th decrease in demand for intermediates ($\beta = 1$), the productivity gap is 50 percen ($s = 0.5$), and labor in the formal sector accepts the same percentage wage cut a labor in the informal sector ($b = 1$), the capital stock still decreases i $\sigma^m < \sigma^x L^x/L^m\theta_T$. As in the condition governing the short-run impact on manu facturing labor demand, the capital stock falls if the elasticity of substitution i the manufacturing sector is not many times larger than the elasticity of substitu tion in the agricultural sector.

The high probability that the capital stock will decrease implies that manufac turing employment is more likely to fall in the long run than in the short run The change in employment in manufacturing across steady states is

$$\frac{\hat{L}^m}{\hat{P}} = \frac{\theta_Z}{\theta_L N}\left[\sigma^m\beta(1 - \theta_K) - \sigma^x \frac{L^x(1 - s)\theta_L}{L^m\theta_T b}\right]. \quad (15)$$

The critical value of σ^m required for L^m to increase is $(1 - \theta_K)^{-1}$ times large than the critical value defined in expression 13. Hence the condition for manu facturing employment to increase is roughly twice as demanding in the long ru (the cost share of capital in the manufacturing sector in developing countries i around 50 percent) as it is in the short run.

The preceding analysis argues that the increase in the price of the intermediate will usually provoke capital decumulation. Figure 1 describes the workings of the adjustment process in this the normal case. The initial equilibrium is (A,B,C). Immediately following the price increase, investment, employment in the high-wage sectors, and the wage rate in the agricultural and manufacturing sectors all decline. As the capital stock decreases over time, employment conditions continue to worsen, and real wages and investment continue to fall.[2] The failure of repeated real wage cuts to forestall further reductions in investment and further decreases in nonagricultural employment reflects an inherent feature of the adjustment process. Falling real wages on the transition path are an *induced* response to weakened labor demand brought on by capital decumulation and consequently do not stimulate employment growth or investment spending.

When the productivity gap (s) is exceedingly large, it is possible that the capital stock will increase, as indicated by the dynamics resulting from the initial equilibrium (D,E,F) in figure 1. After the initial shock, capital accumulation bolsters labor demand in the high-wage sectors, thus driving up the agricultural and manufacturing sector wages. Real output may eventually increase,[3] but the labor market never fully recovers. In the new steady state (X,Y,Z), formal sector employment and real wages are lower.

III. Public Sector Layoffs

Cuts in public sector employment release resources to the private sector. Layoffs connected with higher prices for intermediate inputs are part of a policy package that subjects the private sector to a joint supply shock. By contrast layoffs in those branches of the government that produce final goods and "services" (broadly defined) combine the release of labor resources with a cut in consumption. Layoffs of this type can be analyzed by deleting the intermediate input from the model and letting $Q(L^g)$ represent the value of government services measured in units of tradable goods. Assuming the government cannot charge for its services, a reduction in public sector employment of $-dD/w^g$ maintains fiscal balance when debt service increases.

Initially, the cut in public sector employment increases the supply of labor to

2. The lump-sum tax adjusts to offset the fiscal effects of variations in public sector employment and demand for the intermediate input (Z). Real wages here are thus gross real wages (which differ from net real wages if some part of the tax falls on labor). Since demand for the intermediate input falls with capital stock, the tax is rising on the transition path if and only if $\beta w^g/PZ' < 1$. Net real wages, therefore, fall faster than gross real wages if labor is paid its marginal product in the parastatal sector. In subsequent sections the lump-sum tax is constant, and it is not necessary to distinguish between variations in gross and net real wages.

3. Real output always increases if the government lays off workers as dictated by the drop in demand for the intermediate input and if the marginal product of public sector labor exceeds the agricultural wage.

the agricultural sector, thereby depressing agricultural and manufacturing sector wages. Real output may rise or fall in the short run depending on the productivity of public sector labor and the division of new hires between manufacturing and agriculture.

If there is an initial contractionary phase, it ultimately proves to be temporary. Lower real wages spur greater investment spending, and, as the capital stock grows, employment in manufacturing increases further and the agricultural wage starts rising. Over the long run the capital stock increases enough that all of the laid-off workers are absorbed in the high-wage manufacturing sector without lowering real wages. (In terms of figure 1, point E is horizontally to the left of point Y, and F is vertically below Z.) To establish this result, observe that in long-run equilibrium the capital rental (r) is tied down by the rate of time preference ($r = \rho + \delta$). It then follows from the zero profit conditions that real wages and the land rental are also constant across steady states. Thus employment in agriculture is unchanged at the new long-run equilibrium, and clearing of the labor market implies that the increase in employment in the manufacturing sector is equal to the decrease in the public sector ($dL^m = - dL^g$).

What is appealing in these results is that eventually higher debt service is financed partially or wholly by an expansion in economic capacity. It is, however, a long step from this to the conclusion that public sector layoffs (in the final goods and services sectors) constitute an easy remedy to the debt problem. A potentially difficult intertemporal tradeoff exists when output decreases in the short run. Furthermore, even if layoffs generate a favorable output path, the distributional repercussions may not be judged acceptable. Real wages and formal sector employment are lower everywhere on the transition path until the new steady state is reached. A prolonged bout of greater inequality is the price paid for higher output in the long run.

IV. Reductions in Public Investment

Fiscal belt tightening often takes its greatest toll on public investment. Cutbacks occur not only in planned infrastructure and industrial projects, but also in a wide variety of education, health, and training programs. I investigate below the repercussions of reducing infrastructure investment. A very similar analysis applies, however, to cuts in government expenditures that foster human capital formation (see Buffie forthcoming).

Infrastructure capital serves to enhance the productivity of private capital and labor. The simplest way to capture this complementarity of social infrastructure and private inputs is to introduce urban infrastructure capital as a third distinct factor in the manufacturing sector. The stock of social infrastructure (K^i) is fixed in the short run and rises or falls over time depending on whether net public investment is positive or negative. All rents generated by social infrastructure accrue to private capitalists—there is no charge for the productive services yielded by the stock of social infrastructure.

Consider now what happens in the long run when greater debt service forces a

cut in infrastructure investment (I^i). With factors being complementary, the productivity of labor in private manufacturing declines. Consequently, real wages and employment in private manufacturing fall across steady states.

In view of the countervailing effects exerted by a smaller stock of social infrastructure and a lower manufacturing sector wage, it might appear that the impact on the incentive to accumulate capital is generally ambiguous. This is not the case. The decrease in the wage is an induced, second-round response; as such it is too weak to offset the drop in the productivity of private capital caused by the reduction in the stock of social infrastructure. Disinvestment in social infrastructure thus leads to capital decumulation on a broad front. In the new steady state the capital stock is lower by the amount

$$\hat{K} = \frac{\theta_i(\sigma^x L^x + \sigma^m L^m b\theta_T)}{\sigma^m L^m \theta_T b(1 - \theta_K) + \sigma^x L^x \theta_i} \tag{16}$$

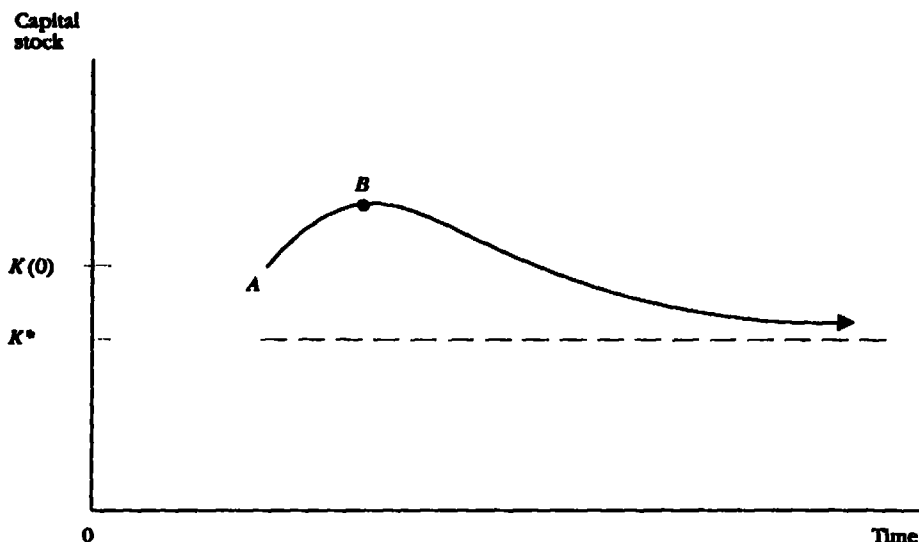
where θ_i is the cost share of social infrastructure. ($\theta_i \equiv r^i K^i / C^m$, where r^i is the implicit rental attached to social infrastructure, and C^m is total costs in the manufacturing sector.)

With two capital stocks varying over time, the dynamics are intrinsically complex, and a variety of adjustment paths are possible. On a "normal" adjustment path, private investment jumps downward on impact but does not overshoot its steady-state level. The two capital stocks, manufacturing sector employment, and real wages all decline monotonically en route to the steady state.

In the normal case the lower equilibrium capital stock elicits an immediate reduction in private investment. There is also, however, the intriguing possibility that investment will increase initially. When the government announces a reduction in infrastructure investment, the representative family firm foresees a declining path for future income and the capital rental. The lower stream of quasi-rents earned by the capital stock implies that eventually investment will decline. But if the family firm has a strong preference for a smooth consumption path, it may increase investment temporarily to shift some consumption from the present to the future. This gives rise to the dynamics shown in figure 2. The possibility of this type of adjustment underscores the importance of bringing the medium and long run into view when evaluating stabilization policy. Over phase *AB*, private and public investment appear to be substitutes, and, *if* the economy experiences a downturn, it is likely to be mild. The short-run response, however, is a faulty guide to how the policy affects the economy's growth prospects. Phase *AB* is only one part of a much longer adjustment process in which public and private capital ultimately prove to be strongly complementary.

V. The Adjustment Process and the Sacrifice Ratio

There are many ways to raise revenue and lower expenditure to achieve a desired reduction in the fiscal deficit. The simple but important point of the analysis in sections II to IV is that the real repercussions of fiscal austerity depend sensitively on how the fiscal deficit is lowered.

Figure 2. *The Adjustment Path of the Capital Stock with an Initial Increase in Investment*

Tables 4 and 5 present evidence on the adjustment costs for public sector price increases and cutbacks in infrastructure investment. These tables show the decrease in net national income (GDP minus external debt service) relative to the increase in debt service. I call this ratio the "sacrifice ratio." In a world in which lump-sum taxes could be employed to service the debt, the sacrifice ratio would equal unity.

In tables 4 and 5 the elasticity of substitution in manufacturing, the degree of labor market dualism (the ratio of the manufacturing wage to the agricultural wage), and the degree of real wage rigidity are allowed to vary. The other parameter values underlying the solution grids for both tables are the elasticity of substitution in agriculture (σ^x) at 0.50, the cost share of land in agriculture (θ_T) at 0.475, the cost share of labor in manufacturing (θ_l) at 0.40, the rate of time preference (ρ) at 0.10, the depreciation rate (δ) at 0.05, and the ratio of agricultural to manufacturing sector output (Q^x/Q^m) at 1.7. In table 4 the cost share of the intermediate input (θ_z) is 0.10, the parameter that defines the government's layoff policy (β) is 1, and the percentage gap between the marginal product of labor in private manufacturing and in the parastatal sector (s) is 0.3. In table 5 the cost share of social infrastructure (θ_i) is 0.2.

The values for the ratio of agricultural to manufacturing sector output and the cost shares are set to yield output and employment shares for the manufacturing sector close or equal to those of the industrial sector in the highly indebted developing countries in 1980, the last year before the debt crisis. The ratio of

Table 4. *The Sacrifice Ratio with Adjustment through an Increase in the Price of the Intermediate Input*

Degree of real wage rigidity (b)	Elasticity of substitution in manufacturing			
	0.5	1.0	1.5	2.0
Ratio of manufacturing wage to agricultural wage = 1.25				
0.50	5.0	4.7	4.3	4.0
0.75	3.5	3.2	2.9	2.6
1.00	2.8	2.5	2.2	1.8
Ratio of manufacturing wage to agricultural wage = 1.50				
0.50	6.8	6.5	6.1	5.8
0.75	4.8	4.4	4.1	3.8
1.00	3.7	3.4	3.1	2.7
Ratio of manufacturing wage to agricultural wage = 1.75				
0.50	8.6	8.3	7.9	7.6
0.75	6.0	5.6	5.3	5.0
1.00	4.6	4.3	4.0	3.6
Ratio of manufacturing wage to agricultural wage = 2.00				
0.50	10.4	10.1	9.7	9.4
0.75	7.2	6.8	6.5	6.2
1.00	5.5	5.2	4.9	4.5

Note: The sacrifice ratio is the decrease in net national income relative to the increase in debt service.
Source: Author's calculations.

Table 5. *The Sacrifice Ratio with Adjustment through an Increase in Investment in Infrastructure*

Degree of real wage rigidity (b)	Elasticity of substitution in manufacturing			
	0.5	1.0	1.5	2.0
Ratio of manufacturing to agricultural wage = 1.25				
0.50	7.4	6.9	6.5	6.2
0.75	7.1	6.5	6.1	5.9
1.00	6.9	6.2	5.9	5.6
Ratio of manufacturing to agricultural wage = 1.50				
0.50	8.2	7.6	7.2	6.9
0.75	7.9	7.2	6.7	6.4
1.00	7.6	6.9	6.4	6.1
Ratio of manufacturing to agricultural wage = 1.75				
0.50	8.8	8.2	7.7	7.4
0.75	8.5	7.7	7.2	6.9
1.00	8.2	7.4	6.8	6.5
Ratio of manufacturing to agricultural wage = 2.00				
0.50	9.2	8.6	8.2	7.8
0.75	8.9	8.2	7.6	7.2
1.00	8.6	7.8	7.2	6.8

Note: The sacrifice ratio is the decrease in net national income relative to the increase in debt service.
Source: Author's calculations.

employment in manufacturing to employment in agriculture varies so as to be consistent with the values chosen for other parameters. (This implies that the distribution parameter changes in the CES *production* functions. *Technology differs across the cells in tables 4 and 5.) The employment ratio L^x/L^m is thus higher the more distorted the labor market, rising from 2.79 when $w^m/w^x = 1.25$ to 4.46 when $w^m/w^x = 2$. At $w^m/w^x = 1.5$, the proportion of the private labor force employed in the manufacturing sector* assumes the same value (23 percent) as the weighted average employment share of the industrial sector in the highly indebted countries in 1980. The share of the manufacturing sector in GDP (37 percent) equals the 1980 output share of the industrial sector in the highly indebted countries. (The figures for the output and employment shares of the industrial sector are from the World Bank [1988].)

In table 4 it is assumed that the productivity gap between parastatal and private manufacturing labor is fairly large ($s = 0.3$) and that the government summons the political will to enforce a tough layoff policy ($\beta = 1$). (Since $\beta = 1$, production of government services does not change, and the results measure the impact on total output, not just private sector value added.) To generate the solution grid in table 5, it was necessary to make some assumption about the return on infrastructure capital. I imposed the condition that the initial direct return (the return ignoring the impact on employment in manufacturing and the labor market distortion) equal the return on private capital.

What stands out in both tables is that the sacrifice ratio is so far above unity. The large values reflect the fact that capital decumulation and greater allocative inefficiency in the labor market substantially increase the costs of adjustment. If the capital stock and employment in manufacturing were unchanged, the sacrifice ratio would be unity in table 4 and three in table 5. The actual values for the sacrifice ratio indicate that the real loss of output attributable to adverse general equilibrium repercussions is usually several times larger than the combined loss owing to higher debt service and the direct contractionary effect produced by the policy shift. It is notable that real wage flexibility helps a good deal in table 4 but not much in table 5. Increasing the percentage wage cut in the formal sector from 50 to 100 percent of the percentage wage cut in the informal sector (that is, $b = 1$ versus $b = 0.5$) seldom lowers the sacrifice ratio for a reduction in infrastructure investment by more than 10 percent.

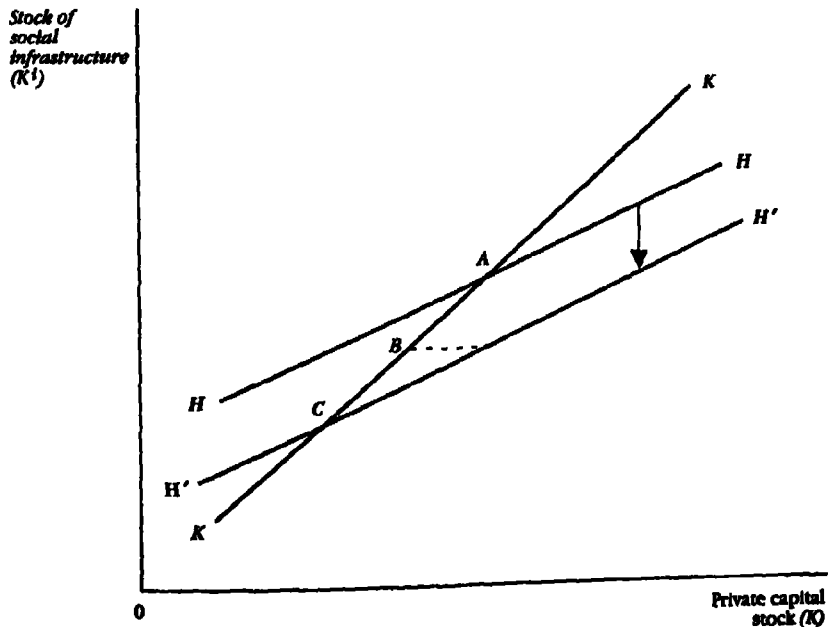
Is it sensible to take seriously the quantitative predictions of a simple model? In this instance I believe the answer is yes. Sacrifice ratios on the order of 5–10 do not seem particularly unrealistic. Since the growth rate is zero across steady states in the model, a sacrifice ratio of eight should be interpreted as saying that the adjustment to an increase in debt service equal to 3 percent of gross national product would entail a cumulative output loss of 24 percent before the economy recovered its previous trend growth rate. This might take the form of the per capita growth rate being (approximately) two points lower for 10 years. In light of the deep contractions suffered by many debtor countries, it is not obvious whether the sacrifice ratios in tables 4 and 5 are too large or too small.

VI. Feedback Effects and the Government Budget Constraint

So far I have assumed that taxes are lump sum. Because of this assumption, the previous analysis, grim as it was, understates the difficulty of adjustment. Under the more realistic specification that tax revenues depend upon output, one-shot fiscal adjustments no longer suffice to meet the debt service target. Any measure that lowers real output lowers tax revenues as well, necessitating further fiscal retrenchment. It is all too easy for the economy to fall into a vicious, contractionary spiral in which capital decumulation, worsening underemployment, and fiscal difficulties become mutually reinforcing.

For illustrative purposes, return to the model of section IV and replace the lump-sum tax by a flat value added or income tax t. As tax revenues are now endogenous, infrastructure investment must be adjusted to satisfy the government budget constraint. This makes public and private investment strongly interdependent. A decrease in the private capital stock leads to a reduction in tax revenues and a matching cut in public investment. The subsequent decrease in the supply of social infrastructure depresses private capital accumulation still more, which results in a further loss of tax revenues, and so forth.

The joint dependence of the two capital stocks (the private capital stock and social infrastructure) is depicted in figure 3. The upward sloping KK schedule is

Figure 3. ***The Joint Dependence of the Stock of Social Infrastructure and the Private Capital Stock***

based on equation 16 and reflects the positive relationship between the equilibrium capital stock and the stock of social infrastructure. The *HH* schedule shows how social infrastructure varies with the capital stock as needed to comply with the government budget constraint. A stable underlying adjustment mechanism requires that, after taking account of the induced reduction in output (for a given private capital stock), the fiscal surplus increases in the long run when social infrastructure is reduced. Given this, *HH* is also positively sloped.

An increase in debt service shifts *HH* vertically downward while leaving the *KK* schedule in place. If taxes were lump sum, point *B* would be the new long-run equilibrium. But when revenues derive from a value added tax, the lower level of output at *B* produces a fiscal deficit. Facing a revenue shortfall, the government further reduces investment in infrastructure, which leads to another round of capital decumulation, lower tax collections, and additional expenditure cuts. If *HH* intersects *KK* from below, private and public capital decumulation feed back upon one another in a destabilizing fashion, and the downward spiral continues until either debt service is suspended or a greater share of the adjustment is shifted onto fiscal instruments less harmful to investment. In figure 3 a stable process operates (*HH* intersects *KK* from above), so the economy eventually converges to point *C*. The distance between *B* and *C* measures the additional capital decumulation (private and public) owing to the feedback effects. This may well exceed capital decumulation directly attributable to higher debt service. It can be shown, for example, that when $2t > \delta r^s$ (where r^s is the implicit rental on infrastructure capital) the respective decreases in the capital stock, social infrastructure, and the agricultural wage are more than twice as large as the decreases that occur under lump-sum taxes.

The message here accords, I believe, with recent macroeconomic history in several developing country debtors. The adjustment to higher debt service is long and traumatic because, once growth decelerates, fiscal problems become nearly systemic and the government finds it is compelled, year after year, to make cuts in productive expenditures. Experience to date and the large sacrifice ratios in tables 4 and 5 suggest that this vicious cycle, if not actually unstable, requires considerable time to work itself out. Macroeconomic austerity can acquire a life of its own.

VII. Summary and Concluding Remarks

One of the fundamental, unresolved puzzles in development macroeconomics concerns why austerity programs have produced such deep and prolonged recessions in many debtor nations. In this article I have argued that part of the answer to the puzzle may lie in the measures directed at lowering the fiscal deficit. There are sound reasons for thinking that fiscal policy exerts a stronger influence on private investment than is commonly believed. Cuts in public infrastructure investment and higher prices for publicly produced intermediates depress private

investment by lowering usage of factors complementary to capital. Once capital decumulation sets in, employment growth slows in the high-wage sectors of the economy. As a result, formal sector employment declines at the same time as real wages are subject to general downward pressure. Moreover the adjustment process is likely to be protracted by ongoing budgetary problems. If slower growth leads to lower tax collections, further fiscal retrenchment becomes necessary as the economy contracts. It may take a long time to escape the strong, interlocking grip of slow growth and chronic fiscal deficits. In fact, as shown in section VI, endogenously driven fiscal austerity may account for the greater part of the losses suffered when adjusting to higher debt service.

Public sector wage cuts and layoffs in the final goods and services sectors offer better prospects for maintaining growth in the face of increased debt service. Wage cuts are a pure absorption-reducing policy that need not have any lasting adverse impact on real output. Layoffs in the final goods and services sectors are actually conducive to adjustment through growth. But although these two policies appear to work better than other fiscal measures, neither can be recommended without qualifications. For public sector layoffs, the adjustment process that brings gains over the long run also entails lower real wages and lower formal sector employment in the short and medium run. Given the valid concern of policymakers to minimize the impact of adjustment policies on the poor, layoffs will often have to be phased in slowly.

Public sector wage cuts are less objectionable on distributional grounds but have other drawbacks. After eight years of adjustment *further* wage cuts may no longer be a real policy option in some highly indebted countries. In several Latin American and African nations real public sector wages have fallen to the point where moonlighting, long lunch breaks, and shirking have severely undermined the government's capacity to carry on normal operations.

There are two broad lessons for policy in this analysis. First and most obvious, something must be done to revive investment. In addition to increased spending on social infrastructure and human capital formation, substantial investment subsidies are needed to overcome the divergence between social and private returns caused by underemployment. In short, a "big push" in *productive* government expenditures is required. This points, of course, to the urgency of tax reform. Without a wider tax base and heavier taxation of factors in inelastic supply, there is little hope the fiscal bind will loosen enough to allow per capita growth rates to again reach respectable levels.

The second broad lesson is that adjustment should be gradual. Large-scale fiscal adjustments are often necessary. But if all adjustment is to take place within just a year or two, it is inevitable that a wide range of contractionary measures will be adopted, including many that are incompatible with policymakers' output and employment targets. With new capital inflows financing a longer timetable for reform, more of the burden of adjustment can be shifted onto the small set of fiscal instruments that do not damage the economy's long-run prospects for development.

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Arida, Persio, and Lance Taylor. 1989. "Short-Run Macroeconomics." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics*, vol. 2. New York: North-Holland.

Arrow, Kenneth, and Mordechai Kurz. 1970. *Public Investment, the Rate of Return and Optimal Fiscal Policy*. Baltimore, Md.: Johns Hopkins University Press.

Boadway, Robin. 1978. "Public Investment Decision Rules in a Neo-Classical Growing Economy." *International Economic Review* 19: 265–87.

Buffie, Edward. Forthcoming. "The Long-Run Consequences of Short-Run Stabilization Policy." In Susan Horton, Ravi Kanbur, and Dipak Mazumdar, eds., *Labor Markets in an Era of Adjustment*. EDI Development Study. Washington, D.C.: The World Bank.

Gindling, Timothy. 1989. "Crisis Economica y Segmentacion del Mercado de Trabajo en el Area Urbana de Costa Rica." *Ciencias Economicas* (Costa Rica) 9 (1–2): 79–93.

Gregory, Peter. 1975. "The Impact of Institutional Factors on Urban Labor Markets." Studies in Employment and Rural Development. Washington, D.C.: The World Bank.

House, William J. 1984. "Nairobi's Informal Sector: Dynamic Entrepreneurs or Surplus Labor?" *Economic Development and Cultural Change* 32: 277–302.

International Monetary Fund. Various years. *World Economic Outlook*. Washington, D.C.

Mazumdar, Deepak. 1976. "The Urban Informal Sector." *World Development* 4: 655–79.

———. 1989a. "Government Intervention and Urban Labor Markets in Developing Countries." EDI Working Paper, World Bank, Washington, D.C. Processed.

———. 1989b. "Microeconomic Issues of Labor Markets in Developing Countries: Analysis and Policy Implications." EDI Seminar Paper 40. World Bank, Washington, D.C. Processed.

Marchand, Maurice, Pierre Pestieau, and John Weymark. 1982. "Discount Rate for Public Enterprises in the Presence of Alternative Financial Constraints." *Journal of Economics* Supplement 2: 27–50.

Merrick, Thomas. 1976. "Employment and Earnings in the Informal Sector in Brazil." *Journal of Developing Areas* 10: 337–54.

Pestieau, Pierre. 1974. "Optimal Taxation and Discount Rate for Public Investment in a Growth Setting." *Journal of Public Economics* 3: 217–35.

———. 1975. "The Role of Taxation in the Determination of the Social Discount Rate." *International Economic Review* 16: 362–68.

Portes, Alejandro, Silvia Blitzer, and John Curtis. 1986. "The Urban Informal Sector in Uruguay: Its Internal Structure, Characteristics and Effects." *World Development* 14: 727–41.

Rader, Trent. 1968. "Normally, Factors Are Never Gross Substitutes." *Journal of Political Economy* 76: 38–43.

Sachs, Jeffrey. 1989. "Introduction." In Jeffrey Sachs, ed., *Developing Country Debt and the World Economy*. University of Chicago Press.

Squire, Lynn. 1981. *Employment Policy in Developing Countries*. New York: Oxford University Press.

United Nations. Various years. *Economic Survey of Latin America and the Caribbean*. Santiago, Chile.

World Bank. 1988. *World Development Report*. New York: Oxford University Press.

Yoshida, Masatushi. 1986. "Public Investment Criterion in an Overlapping Generations Economy." *Economica* 53: 247–63.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 3: 353-398

Measuring the Independence of Central Banks and Its Effect on Policy Outcomes

Alex Cukierman, Steven B. Webb, and Bilin Neyapti

Making the central bank an agency with the mandate and reputation for maintaining price stability is a means by which a government can choose the strength of its commitment to price stability. This article develops four measures of central bank independence and explores their relation with inflation outcomes. An aggregate legal index is developed for four decades in 72 countries. Three indicators of actual independence are developed: the rate of turnover of central bank governors, an index based on a questionnaire answered by specialists in 23 countries, and an aggregation of the legal index and the rate of turnover.

Legal independence is inversely related to inflation in industrial, but not in developing, countries. In developing countries the actual frequency of change of the chief executive officer of the bank is a better proxy for central bank independence. An inflation-based index of overall central bank independence contributes significantly to explaining cross-country variations in the rate of inflation.

"*Willpower is trying hard* not *to do something that you really want to do,*" *said Frog.*

"*You mean like trying* not *to eat all these cookies,*" *asked Toad.*

"*Right,*" *said Frog. He put the cookies in a box.* "*There, now we will not eat any more cookies.*"

"*But we can open the box,*" *said Toad.*

"*That is true,*" *said Frog. He tied some string around the box. He got a ladder and put the box up on a high shelf.* "*There, now we will not eat any more cookies.*"

"*But we can climb the ladder. . . .*" (Lobel 1972)

Institutions cannot absolutely prevent an undesirable outcome, nor ensure a desirable one, but the way that they allocate decisionmaking authority within the public sector makes some policy outcomes more probable and others less likely. An important example of this principle concerns the balance of authority between the central bank and the executive and legislative branches of government. Economists and practitioners in the area of monetary policy generally

Alex Cukierman is with the Department of Economics, Tel Aviv University; Steven B. Webb is with the Country Economics Department at the World Bank; and Bilin Neyapti is with the Country Economics Department at the World Bank and the Economics Department, University of Maryland.

believe that the degree of independence of the central bank from other p
government affects the rates of expansion of money and credit and, th
them, important macroeconomic variables, such as inflation and the size (
budget deficit.

Ultimately, the central bank's authority and scope of action depends or
government. But governments often pass laws and follow customs that g
their central banks authority and autonomy *to pursue price stability,* even wl
it conflicts with *other government objectives. Making the central bank* a
agency with the mandate and reputation for maintaining price stability benefi
the economy and the government itself in various ways. Central bank independence is one of the means by which a government can choose the strength of its commitment to price stability (Cukierman 1992a, chap. 23, and 1992b; Lohmann 1992). A vast literature discusses the costs of inflation (see Fischer *[1986] for a survey)*; the central bank's pursuit of price stability can help reduce these costs. Price stability is also necessary, although far from sufficient, for developing a local capital market where both government and businesses can borrow more conveniently and cheaply in the long run. International financing, such as for the countries recovering from hyperinflation in the 1920s, has often been conditional on the central bank's mandate and authority to pursue the stability of prices and exchange rates.

Pursuing price stability necessarily competes at least some of the time with other tasks that central banks can and often do perform—such as managing the government's financial transactions, financing the government's deficits with money issue, financing development projects, and bailing out insolvent businesses, including banks and publicly owned enterprises. Although most governments recognize the long-run benefit of price stability, other goals often loom larger in the short run. Assuring price stability, therefore, usually requires ensuring that the central bank is not forced to perform these functions, at least not when they would cause inflation. Sometimes the government or the treasury takes direct responsibility for limiting the demands on a subservient central bank. Even in these cases, but especially in the more typical case where the government has strong tendencies to focus on issues other than price stability, central bank independence and an explicit mandate to pursue price stability are generally regarded as important institutional devices for ensuring price stability.

This belief has eluded comprehensive verification in the past because of the difficulties in measuring the autonomy of central banks independently from

Many individuals helped assemble the data for this study. Particular thanks are due to Walter Wasserfallen, Herman-Josef Dudler, John Flemming, D. Hiss, Wolfgang Schill, Marc-Olivier Strauss-Kahn, Ann Johannessen, and those who answered the questionnaire. The authors also received useful comments from James Alt, Tomas Balino, Edgardo Barandiaran, Paul Beckerman, Mario Blejer, Michael Bruno, Gerald Caprio, Max Corden, Andrew Crockett, Jose de Gregorio, Patrick Downes, Robert Effros, Valeriano Garcia, Sergio Pereiro Leite, Alfredo Leone, Klaus Richel, Lawrence Summers, V. Sundararajan, Mark Swinburne, Richard Webb, Eduardo Wiesner, and participants at the November 1991 NBER Conference on Political Economics.

ation of the inflation outcomes. Actual, as opposed to formal, central independence depends not only on the law, but also on many other less-tured factors, such as informal arrangements between the bank and other of government, the quality of the bank's research department, and the nality of key individuals in the bank and the (rest of the) government. ause of the difficulty in quantifying such features in an impartial manner, vious studies developed indexes of central bank independence based mostly legal independence—and only for the industrial countries at that. The multi-ountry studies that attempt to rank independence for a cross-section of coun-tries include Bade and Parkin (1980), Skanland (1984), Parkin (1987), Alesina (1988), Masciandaro and Tabellini (1988), Bodart (1990), Swinburne and Castello-Branco (1991), and Grilli, Masciandaro, and Tabellini (1991), plus Leone's (1991) study of limits on lending in several industrial and developing countries. There is also a large literature of single- or multicountry case studies, which includes Mittra (1977), Schokker (1980), Eizenga (1983), Kearney (1984), Dotsey (1986), Epstein and Schor (1986), Bordes and Strauss-Kahn (1987), Bordo and Redish (1987), Eizenga (1987), Keenan and Mayes (1987), Suzuki (1987), Goodhart (1988), Holtfrerich (1988), Willett (1988), Fazio (1991), Cargill and Hutchison (1990), Mayer (1990), Meltzer (1991), Maxfield (1992), and Volcker, Mancera, and Godeaux (1991).

Indicators based only on the law have two problems. First, the laws are incomplete in that they cannot specify explicitly the limits of authority between the central bank and the political authorities under all contingencies. These voids are filled by tradition at best and by power politics at worst. Second, even when the law is quite explicit, actual practice may deviate from it.

This article develops unified and broadly based measures of central bank independence, uses them to rank central banks by their degree of independence, and explores the relation between their independence and inflation outcomes. The study goes beyond previous work in three dimensions. First, the set of countries is wider, including up to 72 countries (21 industrial countries and 51 developing countries). The wider sample makes it possible to examine whether there are systematic differences in central bank independence between industrial and developing countries. Second, the coverage in time goes back to the 1950s, if the bank existed then. Third, the study uses a wider range of information on central bank independence. The spirit of the law and its application in practice are generally more important than the letter of the law. In addition to coding characteristics of the central bank law, the study looks at the actual frequency of turnover of central bank governors and at the questionnaire responses from specialists on monetary policy in a subsample of 23 countries. The questionnaire was designed to identify divergences between the central bank's charter and actual practice. We use several different indicators of independence because, in addition to the noise that they contain, each indicator captures a somewhat different aspect of independence.

Examining the relation of inflation to alternate indicators for independence

reveals that legal independence is an important determinant of inflation in industrial countries. In developing countries, by contrast, governors' turnover strongly and positively associated with inflation. This finding suggests that the are larger divergences between actual practice and the law in developing than i industrial countries (see, for example, Bodart 1990 and Leone 1991).

I. Measures of the Legal Independence of Central Banks

Legal independence is, of course, an essential component of actual independence, but it is also of interest for several other reasons. First, it indicates what is the degree of independence that legislators meant to confer on the central bank. Second, practically all existing attempts at systematically characterizing central bank independence rely solely on legal aspects of independence (Bade and Parkin 1980; Banaian, Laney, and Willett 1983; Skanland 1984; Parkin 1987; Alesina 1988; Masciandaro and Tabellini 1988; Grilli, Masciandaro, and Tabellini 1991). Establishing comparability with previous studies requires an index of legal independence for our sample of countries.

The laws of central banks differ in their focus, scope, and degree of detail. Many provisions in central bank charters have no direct bearing on the issue of central bank independence. Ranking central bank charters by their degree of legal independence is therefore difficult and inevitably requires subjective judgment.

Coding Legal Independence

Our coding of the legal central bank independence followed two principles. First, we coded only a few narrow but relatively precise legal characteristics. Second, we used only the written information from the charters. Additional information on how the law is applied was deliberately left out, since it is reflected by separate indexes that are discussed in section II. These principles make it possible to rank central banks by their degree of independence in various legal dimensions with relatively few subjective judgments and to focus on concrete details of the law rather than on a broader but more impressionistic view of it.

The legal characteristics of the central bank as stated in its charter are grouped into four clusters of issues:

- The appointment, dismissal, and term of office of the chief executive officer of the bank—usually the governor
- The policy formulation cluster, which concerns the resolution of conflicts between the executive branch and the central bank over monetary policy and the participation of the central bank in the budget process
- The objectives of the central bank
- Limitations on the ability of the central bank to lend to the public sector; such restrictions limit the volume, maturity, interest rates, and conditions

for direct advances and securitized lending from the central bank to the public sector.

The clusters were built up from 16 different legal variables, each coded on a scale of 0 (lowest level of independence) to 1 (highest level of independence). The detailed classification and codings appear in table 1. The codes are set so that a higher number indicates what we expected would lead to a stronger mandate and greater autonomy for the central bank to pursue price stability.

In coding various central banks by the degree of independence within each group of characteristics, the following criteria were used. Central banks in which the legal term of office of the chief executive officer (CEO) is longer and in which the executive branch has little legal authority in appointing or dismissing the governor are classified as more independent in the CEO dimension. By the same logic, central banks with wider authority to formulate monetary policy and to resist the executive branch in cases of conflict are classified as more independent in the policy formulation dimension.

For the objective of the central bank, there are six possible ratings, according to the prominence given to price stability compared with other stated objectives that might conflict with price stability. For instance, when the charter specifies price stability as the main or only goal, the bank is classified as being more independent in this dimension than a central bank with objectives in addition to, but not inconsistent with, price stability, such as stable banking. These banks are, in turn, classified as more independent than banks whose objectives include things like full employment, which might conflict with price stability. The objectives variable is designed to capture the legal mandate of the bank to single-mindedly pursue the objective of price stability. Only the central banks in the Federal Republic of Germany and the Philippines have unequivocal legal mandates for price stability. The objectives variable does not, therefore, reflect the general level of independence from government, in contrast to the CEO and policy formulation variables. In Rogoff's (1985) terminology, the objectives variable measures the strength of the "conservative bias" of the bank's charter.

Similarly, we classify a central bank with tighter limits on its lending to the public sector as more independent to pursue the objective of price stability. These limitations encompass a number of more detailed variables, such as separate limitations on advances and securitized lending and restrictions on maturities and on interest rates. The stricter the limitation, the higher the independence coding given to the bank in that dimension. The comparability of various types of limitations is complicated because the limitations are specified in different ways in different countries. In a few countries limitations on lending are specified in absolute cash amounts and in others as a percentage of central bank liabilities. The limitation is formulated in most cases as a percentage of government's revenues from taxes but in a minority of cases as a percentage of government's expenditures. The "bite" of these limitations obviously depends on the magnitudes of the reference variables. Other things being equal, however, abso-

Table 1. *Variables for Legal Central Bank Independence*

Variable number	Description of variable	Weight	Numerical coding
1	*Chief executive officer (CEO)*	*0.20*	
	a. Term of office		
	Over 8 years		*1.00*
	6 to 8 years		0.75
	5 years		0.50
	4 years		0.25
	Under 4 years or at the discretion of appointer		0.00
	b. Who appoints CEO?		
	Board of central bank		1.00
	A council of the central bank board, executive branch, and legislative branch		0.75
	Legislature		0.50
	Executive collectively (e.g. council of ministers)		0.25
	One or two members of the executive branch		0.00
	c. Dismissal		
	No provision for dismissal		1.00
	Only for reasons not related to policy		0.83
	At the discretion of central bank board		0.67
	At legislature's discretion		0.50
	Unconditional dismissal possible by legislature		0.33
	At executive's discretion		0.17
	Unconditional dismissal possible by executive		0.00
	d. May CEO hold other offices in government?		
	No		1.00
	Only with permission of the executive branch		0.50
	No rule against CEO holding another office		0.00
2	Policy formulation	0.15	
	a. Who formulates monetary policy?		
	Bank alone		1.00
	Bank participates, but has little influence		0.67
	Bank only advises government		0.33
	Bank has no say		0.00
	b. Who has final word in resolution of conflict?[a]		
	The bank, on issues clearly defined in the law as its objectives		1.00
	Government, on policy issues not clearly defined as the bank's goals or in case of conflict within the bank		0.80
	A council of the central bank, executive branch, and legislative branch		0.60
	The legislature, on policy issues		0.40
	The executive branch on policy issues, subject to due process and possible protest by the bank		0.20
	The executive branch has unconditional priority		0.00
	c. Role in the government's budgetary process		
	Central bank active		1.00
	Central bank has no influence		0.00
3	Objectives	0.15	
	Price stability is the major or only objective in the charter, and the central bank has the final word in case of conflict with other government objectives		1.00
	Price stability is the only objective		0.80
	Price stability is one goal, with other compatible objectives, such as a stable banking system		0.60
	Price stability is one goal, with potentially conflicting objectives, such as full employment		0.40

Table 1. *(continued)*

Variable number	Description of variable	Weight	Numerical coding
	No objectives stated in the bank charter		0.20
	Stated objectives do not include price stability		0.00
4	Limitations on lending to the government		
	a. Advances (limitation on nonsecuritized lending)	0.15	
	No advances permitted		1.00
	Advances permitted, but with strict limits (e.g., up to 15 percent of government revenue)		0.67
	Advances permitted, and the limits are loose (e.g., over 15 percent of government revenue)		0.33
	No legal limits on lending		0.00
	b. Securitized lending	0.10	
	Not permitted		1.00
	Permitted, but with strict limits (e.g., up to 15 percent of government revenue)		0.67
	Permitted, and the limits are loose (e.g., over 15 percent of government revenue)		0.33
	No legal limits on lending		0.00
	c. Terms of lending (maturity, interest, amount)	0.10	
	Controlled by the bank		1.00
	Specified by the bank charter		0.67
	Agreed between the central bank and executive		0.33
	Decided by the executive branch alone		0.00
	d. Potential borrowers from the bank	0.05	
	Only the central government		1.00
	All levels of government (state as well as central)		0.67
	Those mentioned above and public enterprises		0.33
	Public and private sector		0.00
	e. Limits on central bank lending defined in	0.025	
	Currency amounts		1.00
	Shares of central bank demand liabilities or capital		0.67
	Shares of government revenue		0.33
	Shares of government expenditures		0.00
	f. Maturity of loans	0.025	
	Within 6 months		1.00
	Within 1 year		0.67
	More than 1 year		0.33
	No mention of maturity in the law		0.00
	g. Interest rates on loans must be	0.025	
	Above minimum rates		1.00
	At market rates		0.75
	Below maximum rates		0.50
	Interest rate is not mentioned		0.25
	No interest on government borrowing from the central bank		0.00
	h. Central bank prohibited from buying or selling government securities in the primary market?	0.025	
	Yes		1.00
	No		0.00

Note: The ranking under each criteria indicates the degree of independence of central banks—the higher the code, the more independent the central bank.

a. Often the law does not contain a separate provision on the resolution of conflict. In those cases, the variable was coded on the basis of the impression from reading the law in its entirety. If the law gives the impression that the government formulates policy guidelines that the bank simply follows, then the ranking is low.

Source: Various central bank laws, Aufricht (1961, 1967); Bank for International Settlements (1963); Effros (1982); and the IMF's computerized files on central bank laws.

lute cash limits are more binding than limits in terms of central bank liabilities, which, in turn, are more binding than limits in terms of government's revenues. The most accommodative limits are those which are specified in terms of government's expenditures. These considerations were embodied in a "type-of-limit" variable and also influenced the classification of the variables for limitations on lending via advances and for limitations on lending via securities. Table 1 shows the details of the several variables for limitations on lending.

Limitations on lending are also classified as stricter the nearer are the rates paid by government to market rates and the shorter are the maturities of the loans from the central bank to the public sector. They are also stricter the narrower the circle of institutions that is allowed to borrow from the central bank and the smaller the discretion of the executive branch to decide to whom and how much the central bank will lend. In addition, central bank laws that prohibit the central bank from buying government securities on the primary market are considered, all things being equal, stricter than laws that do not contain such a prohibition.

The period considered covers the four decades from 1950 to 1989. It is divided into four subperiods: 1950–59, 1960–71, 1972–79, and 1980–89, which we refer to according to the decades in which they are centered. They correspond, respectively, to the gold-dollar standard period before most currencies had convertibility, the period of convertibility with the dollar, the period of the two oil shocks after the end of the Bretton Woods currency system, and the period of disinflation and the debt crisis. Legal variables were coded separately for each decade. Since central bank legislation changes relatively slowly, the codes are, in many cases, identical across subperiods. Nonetheless this procedure captures important legislative changes for some countries. Only one code per decade was assigned for each country for each legal variable. Whenever a change occurred within a decade, the classification was done in line with the legislation that was in effect during at least half of that decade. When a central bank was founded within a decade, its legal variables were coded only if it existed for at least three years during the decade. The coded variables appear in Table A-1 following the concluding section.

Aggregating the Legal Variables

The individual components of legal independence are aggregated in two steps to yield a hierarchy of indexes. Later in the article we investigate the association of those indexes with other variables, such as other indicators of central bank independence and inflation. The basic data on the 16 legal variables described in table 1 were aggregated into eight legal variables as follows. The four variables concerning the appointment and term of office of the governor of the central bank were aggregated into a single variable labeled CEO, equal to the mean of the four components. The three variables under policy formulation were aggregated into a single variable by computing a weighted mean of the variables in that

group, with weights of 0.5 for the resolution of conflict, 0.25 for who formulates monetary policy, and 0.25 for active role of the central bank in formulating the government budget. The objectives variable was treated separately. The first four variables for limits on lending were treated separately; the last four variables in the group were averaged with equal weights into a single variable. This aggregation procedure produces one summary legal variable for each of the first three groups in table 1, and five legal variables for the limitations on lending group. When an entry is not available (—) for one or more variables within a subgroup, only the variables with meaningful entries are aggregated. In such cases the weights of the missing variables are allocated proportionally to the remaining variables within the subgroup.

When the legal variables appear at a high level of disaggregation (as in table 1), a missing observation on at least one variable precludes the use of that country or decade. Partial aggregation alleviates this problem by reducing the number of observations with entries that are not available. In addition, multicollinearity among the 16 legal variables reduces the precision of the estimated effect of each of them on inflation. Partial aggregation alleviates this problem, too.

The eight legal variables from the first round of aggregation were aggregated further into a single index for each country and decade, using weights that we considered most plausible. The weights are indicated in table 1. When a component is missing, the weights of those remaining are expanded proportionately to sum to 1.0. For all the observations, the weights of the component variables summed to at least 0.7. The variable for the legal central bank independence aggregated in this way is similar to a variable aggregated with equal weights; the two are highly correlated.

Table 2 ranks industrial and developing countries according to their aggregate variable for legal central bank independence for 1980–89 and also provides their average inflation rates in the 1980s. Countries classified by the World Bank as low- or middle-income, on the basis of 1985 incomes, are classified as developing; the others are referred to as industrial. Observations for each country and decade are given in appendix table A-1. Austria, Germany, and Switzerland—all industrial countries—have the highest legal independence, while Morocco and Poland—both developing countries—have the lowest. Otherwise the two country groups have very similar distributions of aggregate legal independence. The medians are virtually identical—0.33 and 0.34.

II. Informal Indicators of Actual Independence

The legal status of a central bank is only one of several elements that determine its actual independence. Many central bank laws are highly incomplete and leave a lot of room for interpretation. As a result, factors such as tradition or the personalities of the governor and other high officials of the bank at least partially shape the actual level of central bank independence. Even when the law is

Table 2. *Legal Central Bank Independence and Average Annual Inflation, 1980–89*

Industrial economy			*Developing economy*					
Economy	*Legal central bank independence*[a] *(index)*	*Average annual rate of inflation*[b] *(percent)*	*Economy*	*Legal central bank independence*[a] *(index)*	*Average annual rate of inflation*[b] *(percent)*	*Economy*	*Legal central bank independence*[a] *(index)*	*Average annual rate of inflation*[b] *(percent)*
Germany, Fed. Rep. of	0.69	3	Greece	0.55	18	Botswana	0.33	10
Switzerland	0.64	3	Egypt	0.49	16	Zambia	0.33	25
Austria	0.61	4	Costa Rica	0.47	23	Ghana	0.31	37
Denmark	0.50	7	Chile	0.46	19	Romania	0.30	4
United States	0.48	5	Turkey	0.46	41	Bolivia	0.30	119
Canada	0.45	6	Nicaragua	0.45	128	Western Samoa	0.30	12
Ireland	0.44	9	Malta	0.44	3	China	0.29	8
Netherlands	0.42	3	Tanzania	0.44	27	Singapore	0.29	3
Australia	0.36	8	Kenya	0.44	10	Korea, Republic of	0.27	8
Iceland	0.34	32	Philippines	0.43	13	Indonesia	0.27	9
Luxembourg	0.33	5	Zaire	0.43	45	Colombia	0.27	21
Sweden	0.29	8	Peru	0.43	108	Thailand	0.27	6
Finland	0.28	7	Honduras	0.43	7	South Africa	0.25	14
United Kingdom	0.27	7	Venezuela	0.43	19	Hungary	0.24	9
Italy	0.25	11	Bahamas, The	0.41	6	Uruguay	0.24	45
New Zealand	0.24	12	Portugal	0.41	16	Panama	0.22	3
France	0.24	7	Argentina	0.40	143	Pakistan	0.21	7
Spain	0.23	10	Ethiopia	0.40	4	Brazil	0.21	119
Japan	0.18	3	Lebanon	0.40	—	Taiwan	0.21	5
Norway	0.17	8	Israel	0.39	72	Zimbabwe	0.20	12
Belgium	0.17	5	Barbados	0.38	7	Qatar	0.20	4
			Uganda	0.38	72	Nepal	0.18	10
			Nigeria	0.37	18	Yugoslavia	0.17	73
			Malaysia	0.36	4	Morocco	0.14	7
			Mexico	0.34	50	Poland	0.10	[illegible]
			India	0.34	9			

— Not available.

a. The potential range of the index for legal central bank independence is from zero (min... [illegible]

b. [illegible]

ite explicit, reality may be very different. For example, in Argentina the legal m of office of the governor is four years, but there is also a tradition that the vernor of the central bank offers to resign whenever the government, or even e finance minister, changes. Argentine governors have invariably adhered to is tradition. As a consequence, the average actual term of office of the gover- or was about one year from 1950 to 1989. Obviously the actual independence f the Argentine central bank is substantially lower than the legal indicators nply. It is hard to find systematic indicators of actual independence when it iverges from legal independence, and we do not pretend to resolve this mea- urement issue fully. Here we develop two indicators of actual, as opposed to egal, central bank independence from the actual frequency of change of the governor and from responses to a questionnaire sent to experts on each country.

Turnover of Central Bank Governors

This indicator is based on the presumption that, at least above some threshold, more rapid turnover of central bank governors indicates a lower level of independence. Indeed, more rapid turnover presumably creates dependence. If the political authorities frequently take the opportunity to choose a new governor, they will at least have the opportunity to pick those who will do their will. Frequent turnover may reflect the firing of those who choose to challenge the government. A government would even have some incentive to appoint a governor with a reputation for some independence, thereby gaining a temporary increase in the potential for stimulating output or collecting resources through seigniorage, and then use up his reputation, as happened with Arthur Burns at the Federal Reserve Bank of the United States in the 1970s.

For high turnover rates, the tenure of the central bank governor is shorter than that of the executive branch. This makes the central bank governor susceptible to influence by the executive branch and discourages the governor from trying to implement longer-term policies, especially those that would extend beyond the election cycle. Because in most countries the electoral cycle is at least four years, it is likely that the threshold turnover, above which independence declines seriously, is somewhere between 0.2 and 0.25 changes a year (for an average tenure of four to five years). One would expect that turnovers at the central bank that occur simultaneously with or shortly after changes in the government would indicate lower independence than turnovers that occur at other times. Further work will investigate this issue.

If, however, a governor stays on for several years and perhaps outlasts several heads of government, thus presiding over price stability, the governor's reputation can become strong enough to resist considerable pressure. The government's desire to preserve financial stability can deter it from challenging a well-established central bank governor.

A low turnover does not necessarily imply a high level of central bank independence, however, because a relatively subservient governor may stay in office

Table 3. Turnover Rates of the Central Bank Governor, 1950–89
(average number of changes a year)

Economy	1950–89	1950–59	1960–71	1972–79	1980–89
Industrial economy					
Iceland	0.03	—	0.09	0.00	0.00
Netherlands	0.05	0.00	0.08	0.00	0.10
Denmark	0.05	0.10	0.08	0.00	0.00
Luxembourg	0.08	0.10	0.08	0.13	0.00
Norway	0.08	0.10	0.08	0.00	0.10
Italy	0.08	0.00	0.08	0.25	0.00
United Kingdom	0.10	0.00	0.17	0.13	0.10
Canada	0.10	0.10	0.08	0.13	0.10
Germany, Fed. Rep. of	0.10	0.10	0.08	0.13	0.10
Australia	0.10	0.00	0.08	0.13	0.20
Finland	0.13	0.20	0.08	0.00	0.20
Switzerland	0.13	0.10	0.08	0.13	0.20
Belgium	0.13	0.10	0.08	0.13	0.20
United States	0.13	0.10	0.08	0.25	0.10
Ireland	0.15	0.10	0.17	0.13	0.20
France	0.15	0.00	0.17	0.25	0.20
Sweden	0.15	0.20	0.00	0.38	0.10
New Zealand	0.15	0.00	0.17	0.13	0.30
Austria	0.15	0.10	0.17	0.25	0.10
Japan	0.20	0.20	0.17	0.13	0.30
Spain	0.20	0.20	0.25	0.25	0.10
Developing economy					
Qatar	0.06	—	—	0.14	0.00
South Africa	0.10	0.00	0.17	0.00	0.20
Barbados	0.11	—	—	0.13	0.10
Taiwan	0.13	0.10	0.17	0.00	0.20
Philippines	0.13	0.00	0.25	0.00	0.20
Honduras	0.13	0.11	0.00	0.38	0.10
Tanzania	0.13	—	0.18	0.13	0.10
Malaysia	0.13	—	0.08	0.00	0.20
Israel	0.14	0.20	0.08	0.13	0.20
Zimbabwe	0.15	0.27	0.17	0.13	0.10
Mexico	0.15	0.10	0.08	0.13	0.30
Kenya	0.17	—	0.36	0.00	0.20
Greece	0.18	0.10	0.08	0.38	0.20
Hungary	0.18	0.38	0.17	0.13	0.10
Lebanon	0.19	—	0.24	0.25	0.10

a long time. This is probably true for countries with exceptionally low turnover rates, such as Iceland, Denmark, Norway, and the United Kingdom.

Table 3 presents the average annual turnover rates in the sample countries for 1950–89 and for each decade within that period. These rates are presented separately for developing and industrial countries. Average turnover rates for 1950–89 range from a minimum of 0.034 (one change in 29 years) in Iceland to a maximum of 0.93 (average tenure of about 13 months) in Argentina. Turnover rates in developing countries extend into a range considerably above the highest rates in the industrial countries. The highest average turnover among the indus-

Table 3. *(continued)*

Economy	*1950–89*	*1950–59*	*1960–71*	*1972–79*	*1980–89*
Developing economy (continued)					
Nigeria	0.19	—	0.17	0.25	0.10
Bahamas, The	0.19	—	—	0.18	0.20
Morocco	0.20	—	0.25	0.00	0.20
Ethiopia	0.20	—	0.00	0.50	0.10
Colombia	0.20	0.20	0.25	0.13	0.20
Romania	0.20	0.40	0.08	0.13	0.20
Portugal	0.20	0.20	0.08	0.25	0.30
Thailand	0.20	0.40	0.08	0.25	0.10
Yugoslavia	0.23	0.30	0.17	0.25	0.20
Indonesia	0.23	0.20	0.33	0.13	0.20
Zaire	0.23	—	0.26	0.25	0.20
Nepal	0.24	0.27	0.33	0.25	0.10
Panama	0.24	—	0.56	0.00	0.20
Pakistan	0.25	0.10	0.33	0.25	0.30
Poland	0.28	0.20	0.25	0.13	0.50
Malta	0.28	—	0.27	0.38	0.20
Ghana	0.28	—	0.33	0.25	0.20
Venezuela	0.30	0.20	0.25	0.25	0.50
Egypt	0.31	0.46	0.33	0.13	0.30
India	0.33	0.20	0.33	0.50	0.30
Peru	0.33	0.30	0.33	0.38	0.30
China	0.34	—	—	—	0.30
Uganda	0.34	—	0.36	0.50	0.20
Nicaragua	0.35	—	0.29	0.38	0.40
Singapore	0.37	—	—	0.00	0.60
Zambia	0.38	—	0.38	0.25	0.50
Turkey	0.40	0.30	0.50	0.38	0.40
Botswana	0.41	—	—	0.44	0.40
Korea, Republic of	0.43	0.31	0.67	0.13	0.50
Chile	0.45	0.20	0.33	0.50	0.80
Uruguay	0.48	—	1.03	0.38	0.30
Western Samoa	0.56	—	—	—	0.56
Costa Rica	0.58	0.20	0.83	0.88	0.40
Brazil	0.68	1.01	0.50	0.38	0.80
Argentina	0.93	0.71	1.08	0.88	1.00

— Not available.

Note: Turnover rates were calculated if at least three years of data were available for the decade.

Source: Correspondence with central banks.

trial countries for 1950–89 is 0.2 (or an average tenure of five years) for Spain. More than half of the developing countries have turnover rates exceeding this maximum.

Because average turnover rates for all industrial countries for 1950–89 are less than 0.2, these rates probably do not reveal much about the variations of independence within that group. But because turnover rates vary widely among the developing countries, however, they seem more likely to reveal variations in the independence of those governors.

A governor's legal term of office does not seem to have much effect on the

Table 4. Questionnarie Variables, Weights, and Numerical Coding

Variable number	Variable description	Weight	Numerical coding
1	Tenure of central bank CEO overlap with political authorities	0.10	
	Little overlap		1.0
	Some overlap		0.5
	Substantial overlap		0.0
2	Limitations on lending in practice	0.20	
	Tight		1.00
	Moderately tight		0.66
	Moderately loose		0.33
	Loose or nonexistent		0.00
3	Resolution of conflict	0.10	
	Some clear cases of resolution in favor of bank		1.0
	Resolution in favor of government in all cases		0.0
	All other cases		0.5
4	Financial independence	0.10	
	a. Determination of the central bank's budget		
	Mostly central bank		1.0
	Mixture of bank and executive or legislative branches		0.5
	Mostly executive or legislative branches		0.0
	b. Determination of the salaries of high bank officials and the allocation of bank profits		
	Mostly by bank or fixed by law		1.0
	Mixture of bank and executive or legislative branches		0.5
	Mostly executive or legislative branches		0.0
5	Intermediate policy targets	0.15	
	a. Quantitative monetary stock target		
	Such targets exist; good adherence		1.00
	Such targets exist; mixed adherence		0.66
	Such targets exist; poor adherence		0.33
	No stock targets		0.00
	b. Formal or informal interest rate targets		
	No		1
	Yes		0
6	Actual priority given to price stability	0.15	
	First priority		1.00
	First priority assigned to a fixed exchange rate		0.66
	Price or exchange rate stability are among the bank's objectives, but not first priority		0.33
	No mention of price or exchange rate objectives		0.00
7	Function as a development bank, granting credit at subsidy rates?	0.20	
	No		1.00
	To some extent		0.66
	Yes		0.33
	The central bank heavily involved in granting subsidized credits		0.00

actual turnover. To explore this issue, we regressed actual turnover rates in the four subperiods on the legal terms of office and on decade dummies to control for possible period-specific effects on turnover. The coefficient of the legal term-of-office variable was negative and statistically significant, but the adjusted R^2 was low (0.07), thus indicating that actual turnover is affected by many other factors besides the legal term of office.

Questionnaire on Central Bank Independence

The other group of indicators of central bank independence is based on responses to a questionnaire that was sent to a nonrandom sample of specialists on monetary policy in various central banks. Some questions involve the same issues that underlie the legal variables, but they focus on the practice rather than the law—for example, central bank objectives, their importance in practice, and the strictness of limitations on lending in practice. Some questions refer to additional issues, such as subsidized credits from the bank to the private sector, quantitative targets for the money stock, the determination of the bank budget, and the degree of actual tenure overlap between the governor and high officials in the executive branch. Although the judgments of those responding to the questionnaire are subjective and not entirely uniform, the responses help to identify divergence between actual and legal independence, particularly when the divergence is large.

Answers to the questionnaire sufficed for coding most of the nine questionnaire variables described in table 4 in 23 countries (The codings for each country are given in appendix table A-2). We coded only the parts of the questionnaires that could be translated into clear rankings and for which an adequate share of the questionnaires had responses. Since the questionnaire was worded in the present tense and since policymakers' thinking is dominated by the recent past, the responses are taken to refer to the 1980s.

Variable 1 is designed to reflect the extent to which the terms of office of the governor and of the board of directors are likely to be independent from government. The more the turnover at the central bank coincides with turnover in the government, the less independent the bank is likely to be, and vice versa. Variable 2 reflects the actual limitations on lending in practice and is coded by applying criteria similar to those used to classify the legal limitations on lending described earlier. The lowest level of independence is assigned if there are no limitations on lending or if the government can adjust the limits very easily.

Variable 3 reflects the extent to which conflicts between the government and the central bank are resolved in favor of the latter. Variable 4 captures two aspects of the financial independence of the bank: the determination of its budget and the setting of salaries of its top officials. It is calculated as a simple average of the variables 4a and 4b. The two parts of variable 5 reflect the relative importance of targets for the monetary stock or interest rates—these targets being precommitments agreed to by the bank and government. Money

stock targets would enhance the pursuit of price stability, because the ba could adhere to them in the face of pressure from the government. A target i the nominal interest rate, however, would typically work to limit the ability the bank to respond to upsurges of inflation. Variable 5 is calculated as a simpl average of the variables 5a and 5b. Variable 6 captures more directly the priority assigned to price stability, and variable 7 reflects the extent to which the central bank has the competing objective of providing subsidized credits to encourage development. Further details appear in chapter 19 of Cukierman (1992a).

Aggregating the seven variables gives the results in table 5. The weights used in the aggregation are based on our priors and are shown in table 4; using equal weights gave an almost identical ranking. The indexes are reported only for countries in which the weights of the responses sum up to at least 0.7. Because it is based on subjective evaluations, the questionnaire-based index probably contains more noise than the index of legal central bank independence, but it also probably contains additional pertinent information about actual independence. The main limit of the questionnaire is the small number of countries with responses, but we hope to expand this in the future.

Table 5. *The Questionnaire-based Index of Central Bank Independence and Average Annual Inflation*

Country	*Questionnaire-based index of central bank independence*	*Average annual rate of inflation,[a] 1980–89 (percent)*
Germany, Fed. Rep. of	1.00	3
Costa Rica	0.81	23
Finland	0.78	7
Australia	0.76	8
Italy	0.73	11
Denmark	0.73	7
Bahamas, The	0.71	6
Luxembourg	0.66	5
France	0.65	7
United Kingdom	0.64	7
South Africa	0.64	14
Zaire	0.61	45
Lebanon	0.59	—
Ireland	0.57	9
Barbados	0.54	7
Uganda	0.53	72
Uruguay	0.49	45
Belgium	0.47	5
Turkey	0.44	41
Tanzania	0.38	27
Peru	0.22	108
Yugoslavia	0.17	73
Ethiopia	0.13	4

— Not available.

a. Inflation is computed in logs.

ble 6. *Rank Correlations between Indexes of Legal Central Bank dependence*

Correlation pair	*Country sample*		
	All	*Industrial*	*Developing*
Legal index and rate of turnover			
1950–89	0.000	0.018	0.011
1980–89	0.000	0.065	0.015
Legal and questionnaire-based indexes[a]	0.041	0.334*	0.056
Questionnaire-based index and rate of turnover[a]	0.068	0.050	0.031

* Significant at 10 percent level.
a. Covers 1980–89.

The questionnaires clearly indicate that central banks in developing countries are less independent than those in industrial countries. Only two industrial countries—Ireland and Belgium—are below the median of 0.60, and only four developing countries—The Bahamas, Costa Rica, South Africa, and Zaire—are above it. This contrasts with the findings for legal independence, where the two country groups do not differ widely, but is similar to the finding for turnover.

Relationships between Indexes of Independence

Table 6 shows the rank correlations between indexes of central bank independence: the legal index, the turnover rate, and the questionnaire-based index. None of the various indexes of central bank independence are closely correlated. Only the correlation between the legal independence index and the questionnaire-based index of independence for the industrial countries is even marginally significant, which suggests that the law is a more important determinant of actual independence in the industrial countries. Since the correlation across these indexes is not high, they can be usefully combined to obtain a better measure of overall central bank independence, which is done toward the end of the article.

III. Inflation and Central Bank Independence

Do countries with more independent central banks have lower rates of inflation? The hypothesis that inflation should be negatively related to the legal and questionnaire variables has two bases. (Recall that for both variables a higher code in the range between 0 and 1 reflects a higher level of independence.) First, there is a presumption that central banks are more concerned about price stability than the political authorities (see Rogoff [1985], for instance). Because actual policy is normally the outcome of a compromise between the central bank and the executive branch, a more independent central bank will have a stronger impact on actual policy, and therefore average inflation will be lower (Alesina and Tabellini 1987; Cukierman 1992a, chap. 18). Second, the legal indepen-

dence variable is intentionally structured to reflect, among other things, the extent to which the central bank has an explicit mandate to pursue price stability at the expense of other objectives. For a given level of independence from the political authorities, a more focused legal mandate to pursue price stability is expected to result in a lower rate of inflation.

The legal independence of the central bank is neither a necessary nor a sufficient condition for low inflation, although, other things being equal, less legal independence contributes to higher inflation. Some of the countries with the highest average rates of inflation, such as Argentina, Peru, and Nicaragua, have rankings of legal independence above the median. However, countries such as Belgium, Japan, Morocco, and Qatar, with very low rates of inflation, are ranked in the lowest quartile of legal central bank independence.

To investigate systematically the relation between central bank independence and inflation, we regressed inflation on the various indexes of central bank independence. Because a higher numerical code assigned to the legal and questionnaire variables indicates a higher level of independence, the hypothesis implies that the effect of each of these variables on inflation is negative. The effect of the turnover of central bank governors, at least above some threshold, is predicted to be positive.

The inflation variable was transformed in order to reduce heteroskedasticity of the error and thus improve the efficiency of the estimate. Most countries had average inflation rates of 20 percent or less, but a few had three-digit inflation rates in some decades. Using the straight inflation rate would give undue weight to these outlier observations. So we transformed each year's inflation rate into inflation divided by one plus the inflation rate and then took the geometric average for the decade. This variable represents the annual real depreciation of a given amount of money; we call it D:

$$D = \pi/(1 + \pi) \quad (1)$$

where π is the inflation rate and D (hence, the transformed inflation rate) takes a value from 0 to 1.0. When inflation is 100 percent a year, D is 0.5.

Table 7 presents regressions of the transformed inflation rate (D) on disaggregated indexes of legal central bank independence, along with the governor's turnover variable. Each observation pertains to one decade in one country. Not all countries have observations for all four decades, because some countries or central banks start after 1950. Inflation was counted only for years when the central bank existed, if it started late in the decade. The results show the importance of the turnover rate for explaining variations of inflation in the whole sample and among the developing countries. None of the disaggregated legal variables has a significant coefficient (at the 5 percent level), and an F-test reveals that the variables as a group are not quite significant at the 10 percent level even within the group of industrial countries.

To overcome the collinearity among the disaggregated legal variables, we ran the regressions with the aggregate index of legal independence. Table 8 reports

Table 7. *The Transformed Inflation Rate, Disaggregated Variables of Legal Central Bank Independence, and the Turnover Rate, 1950–89*

Explanatory variable	All countries	Industrial countries	Developing countries
Intercept	0.09**	0.09***	0.09*
	(2.47)	(3.50)	(1.71)
CEO	−0.00	0.02	0.01
	(−0.10)	(0.54)	(0.18)
Policy formulation	0.05	−0.02	0.09
	(0.90)	(−0.54)	(1.08)
Central bank objectives	−0.04	0.01	−0.08
	(−1.29)	(0.42)	(−1.61)
Limitations on lending			
a. Advances	−0.04	−0.02	−0.04
	(−1.11)	(−0.88)	(−0.72)
b. Securitized lending	0.03	−0.01	0.04
	(0.73)	(−0.13)	(0.69)
c. Terms of lending	0.06	0.01	0.08
	(1.27)	(0.42)	(1.15)
d. Potential borrowers	0.02	0.00	0.03
	(1.15)	(0.33)	(0.90)
e. Others	−0.07	−0.06*	−0.05
	(−1.14)	(−1.79)	(−0.57)
Rate of turnover of central bank governor	0.30***	−0.07	0.30***
	(5.99)	(−1.13)	(4.47)
Dummy: 1950–59	−0.08***	−0.03**	−0.10**
	(−2.92)	(−2.05)	(−2.14)
Dummy: 1960–71	−0.09***	−0.02	−0.12***
	(−4.04)	(−1.48)	(−3.57)
Dummy: 1972–79	−0.02	0.03**	−0.03
	(−1.10)	(2.10)	(−1.18)
$\bar{R}^2$	0.29	0.29	0.27
F-statistic for legal variables[a]	1.24	1.62	1.20
	(0.28)	(0.15)	(0.30)
Number of observations	177	60	117

Note: The dependent variable is the transformed inflation rate, *D*. The *t*-statistics are reported in parentheses under estimated coefficients. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

a. The significance levels are in parentheses.

the results. (We used the index based on our priors for weights. Using the index with equal weights produced similar results.) The key results come out when we split the sample into industrial and developing countries.

For the industrial countries, the aggregate legal variable has a statistically significant coefficient with the predicted negative sign. Laws do make a difference. The turnover rate, always low in any case for this subsample, has a negative sign, contrary to our prediction; the *t*-statistic indicates marginal statistical significance. The most anomalous case is Iceland, with the highest inflation rate and lowest turnover among industrial countries. Dropping Iceland from the industrial country subsample makes the estimated coefficient on the turnover rate slightly positive and totally insignificant. The coefficient on the aggregate index of legal central bank independence becomes more significant (the *t*-statis-

Table 8. *The Transformed Inflation Rate, Aggregate Index of Legal Central Bank Independence, and the Turnover Rate, 1950–89*

Explanatory variable	*All countries*	*Industrial countries*	*Developing countries*	*All countries with decomposed turnover variable*
Intercept	0.09***	0.09***	0.11**	0.10***
	(3.55)	(7.17)	(2.51)	(3.54)
Legal central bank independence (aggregate index)	−0.02	−0.06**	0.01	−0.03
	(−0.39)	(−2.54)	(0.11)	(−0.45)
Rate of turnover of central bank governor	0.28***	−0.08*	0.28***	
	(6.64)	(−1.81)	(4.80)	
Decomposed turnover[a]				
High turnover range				0.27***
				(6.27)
Low turnover range				0.20*
				(1.86)
Dummy: 1950–59	−0.08***	−0.03***	−0.11***	−0.08***
	(−3.31)	(−2.94)	(−2.62)	(−3.33)
Dummy: 1960–71	−0.09***	−0.02**	−0.13***	−0.09***
	(−4.45)	(−2.11)	(−4.14)	(−4.47)
Dummy: 1972–79	−0.02	0.03***	−0.04	−0.02
	(−0.88)	(2.90)	(−1.28)	(−0.86)
$\bar{R}^2$	0.26	0.34	0.23	0.25
Number of observations	214	79	135	214

Note: The dependent variable is the transformed inflation rate, *D*. The *t*-statistics are reported in parentheses under estimated coefficients. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

a. The rate of turnover is in the high range if there are 0.25 or more turnovers per year; it is in the low range if there are fewer than 0.25.

tic is −4.32), and the adjusted R^2 increases to 0.61. Italy also has low turnover rates but high inflation, while Japan has high turnover compared with other industrial countries but low inflation. In Japan, the Ministry of Finance has an unusually strong anti-inflation attitude, as well as strong influence over the central bank.

To explore further the composition of the legal variable, we made an index—the lending-limit index—from only the five components pertaining to limits on lending. These components drive the result for the industrial countries. The *t*-statistic for the lending-limit index is about the same as for the whole composite legal variable. The other components of the legal independence variable—CEO, policy formation, and objectives—do not make any significant contribution to explaining inflation. The lending-limit index is not significant for the developing countries.

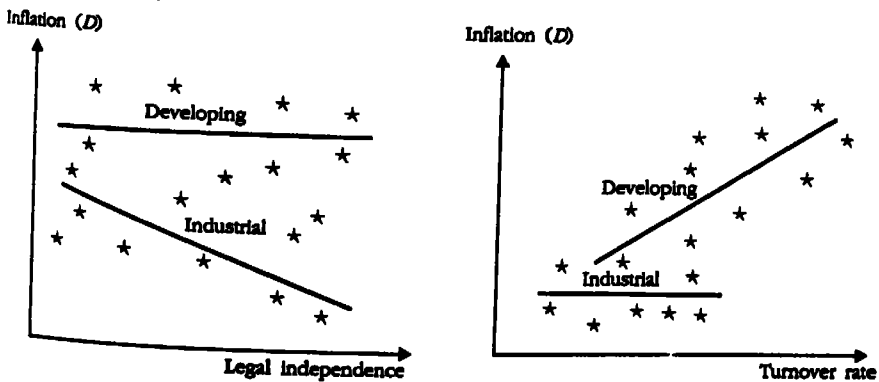
For the developing countries, the turnover rate is highly significant with the predicted positive coefficient. But the aggregate legal variable remains insignificant. This is not to deny that the legal charter has helped ensure the central

bank's independence and commitment to price stability in some developing countries. But statistical evidence for developing countries does not reveal that the central bank laws contribute to explaining the variation of inflation across periods and between countries.

What is the relation between the results for the whole sample and the results for the two subsamples (industrial and developing countries)? In particular, why does the result of the importance of the legal variable for the industrial countries not show up in the whole sample, while the result of the importance of the turnover variable for the developing countries does show up for the entire sample? Figure 1 illustrates these relations.

The left panel in figure 1 shows the partial relation of the aggregate index of legal central bank independence with the transformed inflation rate (*D*). The downward-sloping line shows the significance of the aggregate index of legal central bank independence for industrial countries. The horizontal line shows that the index does not on average affect inflation (*D*) in developing countries. Combining the two samples masks the effect of legal central bank independence on inflation in industrial countries, because including the developing countries raises the average inflation rate (*D*) across the whole range of values for the aggregate index of legal central bank independence.

The right panel shows the partial relation between the rate of turnover of the central bank governor and inflation (*D*). The upward-sloping line shows that the increasing rate of turnover leads to increased inflation in developing countries. The short and low horizontal line shows that the rate of turnover and inflation are both generally lower in the industrial countries. The overall regressions on the rate of turnover reveal a stronger effect of turnover on real depreciation of money in the entire sample than in the developing-country subsample, because

Figure 1. ***Partial Relation of Inflation to the Legal Independence and to the Turnover Rate of Governors of Central Banks in Industrial and Developing Countries***

the overall regression also reflects differences between industrial and developing countries in the general levels of D and of turnover rates.

To explore further the possibility that variations in the rate of turnover of the central bank governor at high rates of turnover might have a stronger effect on inflation than variations at low levels, the regression in the fourth column of table 8 decomposes the turnover variable into high and low turnover ranges. The cutoff of 0.25 turnovers a year or an average tenure of four years seems reasonable, in view of the typical length of electoral cycles. The coefficient on turnover for the high range is very significant, although it is significant for the low range at only the 10 percent level. This result is consistent with the view that turnover is more negatively associated with actual central bank independence in the high range for turnovers.

Although the observations of the questionnaire results (and inflation) were available for a much smaller sample—only 22 countries and one period, the 1980s—than the sample for the turnover and aggregate legal central bank independence variables, we combined it into the regression reported in column 1 of table 8 and obtained the following results:

$$(2) \quad D = 0.27 + \underset{(0.62)}{0.16}\ \text{Legal-CBI} + \underset{(2.22)^{**}}{0.57}\ \text{TOR} - \underset{(-2.97)^{***}}{0.46}\ \text{Question-CBI}$$

$$\text{adjusted } R^2 = 0.38,\ N = 22$$

where ** and *** indicate significance at the 0.05 and 0.01 levels, respectively, Legal-CBI is the index of aggregate legal central bank independence, TOR is the rate of turnover of the central bank governor, and Question-CBI is the aggregate questionnaire-based index computed with the weights listed in table 4. The questionnaire variable has the predicted sign and is very significant statistically. The turnover variable also remains significant, which suggests that it reveals information about actual central bank independence that is not captured by the questionnaire. (The coefficient on the turnover variable may also reflect some simultaneity with the inflation variable, which is discussed below.)

To test the robustness of our results, we tried several variants of the regressions in table 8. Using regular inflation rates instead of the transformed version (D) as the dependent variable yields qualitatively similar results. Using versions of the legal central bank independence and questionnaire variables that were aggregated using equal weights on the component variables did not substantially change the results. We also tried aggregating legal variables with weights derived from principal components analysis. For the industrial-country subsample, the first principal component gave results very similar to those with equal weights and with the weights we selected. The aggregations with weights from the second and third principal components did not have significant coefficients. For the overall sample and the developing-country subsample, the aggregation with weights from the first principal component was not statistically significant.

The turnover rate also reflects the extent to which the government complies

with the law's specification of the governor's term of office. To address this issue more directly, we generated a variable (the compliance variable) equal to the ratio of the actual average term in office to the legal term of office in each country and decade. In most countries and decades, actual average terms in office (including reappointments) are shorter than the legal term. Our hypothesis is that the lower the actual tenure compared with the one in the law, the lower the actual independence of the central bank and the higher the inflation rate (*D*). When the compliance variable is entered instead of the turnover variable, it gives qualitatively similar results to those with the turnover variable—significant coefficients for the overall sample and the developing countries but not in the group of industrial countries. When the compliance and turnover variables are entered together on the right-hand side, however, the compliance variable loses all significance, while the coefficient on the turnover variable remains positive and significant. In other words, it seems to be turnover itself that affects the ability and will to control inflation, rather than the relation of actual to legally stipulated turnover.

We also explored whether the low turnover would identify a subset of developing countries where the degree of legal independence did contribute to explaining inflation. An interaction term of the aggregate index of legal independence times a dummy for low turnover (0.25 or below) was not significant, however, for the whole sample or for the developing countries. Thus low turnover does not seem to be sufficient to reveal or engender a systematic respect for the legal stipulations of independence.

Previous work on inflation and central bank independence is based only on legal data and, in some cases, refers to a subset of the developed countries (Alesina 1988; Grilli, Masciandaro, and Tabellini 1991). These analyses find a significant negative relationship between some of their indexes of central bank independence and average inflation in some periods. Because the indexes of independence used here differ from those used in the previous studies, it is instructive to reexamine their samples with our indexes. We did this with the country subsamples from the earlier studies, and our variables give a better fit with the country subsamples from the previous studies than with the whole industrial-country subsample. The Grilli and others (1991) sample includes Greece and Portugal, which are in our developing-country subsample. (Further details appear in Cukierman [1992a], chap. 20.)

The same degree of central bank independence may be associated with different rates of inflation when there are different economic shocks. It is also possible that the degree of independence affects the response of policy to shocks. A more dependent central bank, for example, may inflate at a higher rate in response to a slowdown in economic activity. The investigation of such interactions is beyond the scope of this article, but it represents an intriguing avenue for future empirical research.

In summary, legal independence is systematically and inversely related to

inflation in industrial, but not in developing, countries. In the latter group the actual frequency of change of the CEO *of the bank is a better proxy for central bank independence.* The divergence between the letter of the law and actual practice seems substantially higher in developing than in industrial countries. This may be due to a general norm of more adherence to the law in industrial countries.

Two-Way Causality between Inflation and Central Bank Independence

There may be two-way causality between inflation and the actual degree of central bank independence. As shown above, less independence contributes to higher inflation. However, high inflation is likely to result, at least after a while, in less independence. High inflation encourages processes that make it easier for the government to influence monetary policy even if the bank charter does not change. Most central bank laws are highly incomplete contracts that do not fully delimit the areas of responsibility of the bank and of the executive branch. In times of high inflation it is harder for the bank to closely control the money supply. In addition, high inflation is partly blamed on the bank, which tarnishes its public image and thus reduces its authority in relation to the treasury, even if this contradicts the charter. Argentina is a dramatic example. If inflation affects actual independence, we expect it also affects the rate of turnover of governors. Therefore, at least part of the positive relation found between inflation and turnover may reflect the effect of inflation on the independence of central banks and therefore on the turnover rate of their CEOs. The legal charters change infrequently and do not seem likely to be simultaneously determined with inflation outcomes.

To examine the possibility that the results for the turnover variable in tables 7 and 8 result from a simultaneity bias, we reestimated the regressions using two-stage least squares and introduced an instrumental variable for the rate of turnover of the central bank governor. The instruments used are transformed inflation (D) and the turnover variable in the previous decade, legal central bank independence, the legal term of office, and the decade dummies. Because periods are about a decade each, the previous observation is about 10 years away from the center of the current one—enough to be predetermined. The use of lagged values of D and turnover as instruments cuts the number of observations used in the equation by about a quarter, to 142. The coefficient of turnover remains positive, actually increases, and remains highly significant (its t-statistic is 6.98). Qualitatively similar results were also obtained with the developing-country subsample.

To investigate further the causality between inflation and the turnover of central bank governors, we did a simple Granger causality test by estimating the bivariate autoregressive processes for inflation and turnover. The periods are the four approximate decades used throughout the article. The long periods seem appropriate for slow-moving processes such as the erosion or the buildup of central bank independence and its interaction with inflation. The estimated

processes for transformed inflation (D) and turnover (TOR) are given in equations 3 and 4 with t-statistics in parentheses under the coefficients.

$$D = 0.02 + \underset{(7.54)^{***}}{0.79\, D_{-1}} + \underset{(3.53)^{***}}{0.17\, \text{TOR}_{-1}} \tag{3}$$

$$\text{TOR} = 0.09 + \underset{(3.41)^{***}}{0.54\, D_{-1}} + \underset{(5.93)^{***}}{0.43\, \text{TOR}_{-1}}. \tag{4}$$

The coefficient of lagged turnover in the inflation equation is highly significant, as is the coefficient of lagged inflation in the turnover equation. A similar picture emerges when the straight inflation rate is substituted for D. The pattern also holds up for the developing-country subsample, but not for the industrial-country subsample, as the earlier results would lead one to expect.

These results imply that there is a vicious circle between inflation and low levels of central bank independence. When sufficiently sustained, inflation erodes central bank independence. Then low independence contributes to higher inflation. Although central bank independence and price stability reinforce each other, the significant coefficients for the turnover variable reported in tables 7 and 8 do seem to reflect a true effect of central bank independence on inflation, and not just simultaneity.

Central Bank Independence and the Variability of Inflation

Variability of inflation imposes economic costs. Indeed, many of the costs of high inflation arise because it is usually more variable and uncertain when the average is high (Cukierman 1984, chap. 3–6). Furthermore, theory implies that inflation will be more variable when the central bank is less independent (Cukierman 1992a, chap. 18). Thus it is important to investigate whether our proxies for central bank independence actually affect the variability of inflation. As a measure of inflation variability, we take the standard deviation of D.

The proxies for central bank independence explain inflation variability to about the same extent that they explain inflation levels (table 9). Turnover contributes significantly to explaining the variance of D in the overall sample and in the developing countries, but the variable for legal independence does not. For the industrial-country sample, the legal variable is significant at the 10 percent level. Again, with Iceland removed from the industrial subsample, the coefficient on turnover is much smaller and more insignificant, and the coefficient for legal central bank independence is significant at the 5 percent level.

The conceptual framework in chapter 18 of Cukierman (1992a) implies that the mean level of inflation and its standard deviation are both negatively related to the degree of independence of central banks across countries. As a consequence the mean level of inflation and its standard deviation are positively related to each other. There is empirical evidence that the variation in the independence of central banks explains some of the correlation between the average level of inflation rates and their variance (Cukierman 1992a, chap. 22).

Table 9. The Standard Deviation of Transformed Inflation, Aggregate Index of Legal Central Bank Independence, and the Turnover Rate, 1950–89

Explanatory variable	All countries	Industrial countries	Developing countries	All countries with decomposed turnover variable
Intercept	0.04***	0.04***	0.05***	0.04***
	(4.16)	(7.42)	(2.88)	(3.96)
Legal central bank independence (aggregate index)	−0.00	−0.02*	0.02	−0.01
	(−0.20)	(−1.77)	(0.40)	(−0.23)
Rate of turnover of central bank governor	0.10***	−0.02	0.09***	
	(6.23)	(−1.28)	(4.05)	
Decomposed turnover[a]				
High turnover range				0.10***
				(5.92)
Low turnover range				0.08**
				(1.95)
Dummy: 1950–59	−0.01	0.00	−0.03	−0.01
	(−1.46)	(0.80)	(−1.52)	(−1.48)
Dummy: 1960–71	−0.03***	−0.01***	−0.04***	−0.03***
	(−3.87)	(−3.14)	(−3.22)	(−3.88)
Dummy: 1972–79	−0.01	−0.00	−0.01	−0.01
	(−1.29)	(−0.77)	(−1.13)	(−1.27)
R^2	0.19	0.18	0.15	0.19
Number of observations	215	79	136	215

Note: The dependent variable is the standard deviation of transformed inflation (*D*). The *t*-statistics are reported in parentheses under estimated coefficients. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

a. The rate of turnover is in the high range if there are 0.25 or more turnovers per year; it is in the low range if there are fewer than 0.25.

Central Bank Independence and the Growth of Credit to the Government

Providing credit to the government would seem to be the most important channel through which the lack of central bank independence leads to inflation because the issue of how to finance its budget deficit is immediately relevant to the government. Providing credit to private and publicly owned business also contributes to inflation, however, as do problems with managing exchange rates. Political authorities are very concerned with these issues as they affect unemployment, bankruptcy rates, export incentives, and the domestic currency cost of imports. Although a systematic evaluation of these considerations lies beyond the scope of this article, we can at least examine the relation between central bank independence and its extension of credit to the government.

Regressing the rate of growth of credit from the bank to the public sector on the two main indicators of central bank independence—the legal independence of the central bank and turnover variables—reveals a pattern similar to, but generally less strong than, what was found with the regressions for the transformed inflation rate (table 10). The aggregate index of legal variables was not significant at all, even in the industrial-country subsample. Turnover of the central bank governor contributes significantly to explaining credit growth to the public sector, although variation of turnover at low rates does not matter much. Within the subsamples of country groups, the coefficients are not statis-

tically significant. Yet they are consistent with the following two hypotheses. The first is that among industrial countries more legal independence limits credit expansion to the public sector, whereas the turnover rate is too low to matter much. The second is that among developing countries higher turnover reflects lower independence, which contributes to faster credit expansion, but the law does not matter much.

That the results for the growth of central bank credit are weaker than the results for inflation suggests that issues other than deficit financing are more important than we had originally thought.

An Overall Index of Inflation-Based Central Bank Independence

Legal independence and turnover capture different dimensions of central bank independence, and each seems to be important for a different subset of countries. This section combines the indicators with a weighting scheme in order to obtain an overall measure of central bank independence. Such weighting is perforce arbitrary, but we reduce the arbitrariness by setting the weights equal to the coefficients from the regressions in which they are used to explain the variation in the transformed inflation variable (D). In this sense, the resulting index of overall independence is based on inflation.

Different regressions for the industrial and developing countries generate the measures of overall central bank independence for members of each group. For

Table 10. *The Rate of Growth of Central Bank Credit to the Public Sector, 1950–89*

Explanatory variable	*All countries*	*Industrial countries*	*Developing countries*	*All countries with decomposed turnover variable*
Intercept	0.22*** (3.06)	0.14** (2.15)	0.27** (2.34)	0.22*** (2.71)
Legal central bank independence (aggregate index)	−0.05 (−0.33)	−0.13 (−1.21)	0.20 (0.67)	−0.05 (−0.33)
Rate of turnover of central bank governor	0.46*** (3.48)	−0.06 (−0.27)	0.27 (1.52)	
Decomposed turnover[a]				
High turnover range				0.46*** (3.38)
Low turnover range				0.45 (1.46)
Dummy: 1950–59	−0.19*** (−2.80)	−0.09* (−1.67)	−0.20* (−1.79)	−0.19*** (−2.79)
Dummy: 1960–71	−0.15*** (−2.73)	−0.01 (−0.28)	−0.22*** (−2.66)	−0.15*** (−2.71)
Dummy: 1972–79	−0.01 (−0.16)	0.10** (2.12)	−0.06 (−0.83)	−0.01 (−0.16)
R^2	0.13	0.14	0.06	0.12
Number of observations	175	68	107	175

Note: The dependent variable is the rate of growth of central bank credit to the public sector. The t-statistics are reported in parentheses under estimated coefficients. * indicates significance at the 10 percent level, ** at the 5 percent level, and *** at the 1 percent level.

a. The rate of turnover is in the high range if there are 0.25 or more turnovers per year; it is in the low range if there are fewer than 0.25.

the industrial countries, the equation has only the aggregate index of le independence, because the coefficient on turnover in tables 7 and 8 had a si contrary to the prediction and because virtually all the observations from t industrial countries had turnover rates below the relevant threshold. For deve oping countries, the equation has both turnover and the index of legal indepe dence. The two equations have the same left-side variable, D, and so the pr dicted values of D offer an inflation-based indicator or index of central ban independence that is comparable across the two subsamples. We estimated th equations across all periods. Table 11 shows the countries ranked by thei

Table 11. *Ranking of Central Banks by an Overall Index of Independence during the 1980s*

Economy	*Transformed inflation rate (D)*		*Legal central bank independence (aggregate index)*	*Rate of turnover of central bank governor*
	Fitted	*Actual*		
Germany, Fed. Rep. of	0.04	0.02	0.69	n.a.
Switzerland	0.05	0.03	0.64	n.a.
Austria	0.05	0.03	0.61	n.a.
Denmark	0.05	0.05	0.50	n.a.
United States	0.05	0.04	0.48	n.a.
Canada	0.06	0.05	0.45	n.a.
Ireland	0.06	0.07	0.44	n.a.
Netherlands	0.06	0.02	0.42	n.a.
Australia	0.06	0.07	0.36	n.a.
Iceland	0.06	0.24	0.34	n.a.
Luxembourg	0.06	0.04	0.33	n.a.
Sweden	0.06	0.06	0.29	n.a.
Finland	0.07	0.06	0.28	n.a.
United Kingdom	0.07	0.05	0.27	n.a.
Italy	0.07	0.08	0.25	n.a.
New Zealand	0.07	0.09	0.24	n.a.
France	0.07	0.06	0.24	n.a.
Spain	0.07	0.08	0.23	n.a.
Japan	0.07	0.02	0.18	n.a.
Norway	0.07	0.07	0.17	n.a.
Belgium	0.07	0.04	0.17	n.a.
Qatar	0.11	0.03	0.20	0.00
Nepal	0.14	0.08	0.18	0.10
Zimbabwe	0.14	0.11	0.20	0.10
Hungary	0.14	0.07	0.24	0.10
Thailand	0.14	0.04	0.27	0.10
Nigeria	0.14	0.16	0.37	0.10
Barbados	0.14	0.05	0.38	0.10
Lebanon	0.14	—	0.40	0.10
Ethiopia	0.14	0.04	0.40	0.10
Honduras	0.14	0.05	0.43	0.10
Tanzania	0.14	0.21	0.44	0.10
Morocco	0.16	0.06	0.14	0.20
Yugoslavia	0.17	0.51	0.17	0.20
Taiwan	0.17	0.03	0.21	0.20

indexes of central bank independence during the 1980s. The actual values of D in the 1980s are presented too, for comparison, along with the legal independence variable for all countries and the turnover variable for the developing countries.

The overall index usually classifies extreme cases in the expected ranges. Thus Germany and the United States are classified near the top and Argentina, Brazil, and Venezuela toward the bottom. For 24 out of the 71 countries (with inflation data) in table 11, the distance between the actual and the predicted transformed inflation rate (D) is less than or equal to 0.03.

Table 11. *(continued)*

Economy	*Transformed inflation rate (D)*		*Legal central bank independence (aggregate index)*	*Rate of turnover of central bank governor*
	Fitted	*Actual*		
Panama	0.17	0.02	0.22	0.20
South Africa	0.17	0.12	0.25	0.20
Colombia	0.17	0.17	0.27	0.20
Indonesia	0.17	0.07	0.27	0.20
Romania	0.17	—	0.30	0.20
Ghana	0.17	0.28	0.31	0.20
Malaysia	0.17	0.03	0.36	0.20
Uganda	0.17	0.47	0.38	0.20
Israel	0.17	0.47	0.39	0.20
Bahamas, The	0.17	0.05	0.41	0.20
Zaire	0.17	0.34	0.43	0.20
Philippines	0.17	0.11	0.43	0.20
Kenya	0.17	0.09	0.44	0.20
Malta	0.17	0.02	0.44	0.20
Greece	0.17	0.14	0.55	0.20
Pakistan	0.19	0.06	0.21	0.30
Uruguay	0.19	0.33	0.24	0.30
China	0.19	0.07	0.29	0.30
India	0.19	0.07	0.34	0.30
Mexico	0.19	0.38	0.34	0.30
Portugal	0.20	0.14	0.41	0.30
Peru	0.20	0.64	0.43	0.30
Egypt	0.20	0.13	0.49	0.30
Botswana	0.22	0.09	0.33	0.40
Nicaragua	0.22	0.67	0.45	0.40
Turkey	0.22	0.28	0.46	0.40
Costa Rica	0.22	0.19	0.47	0.40
Poland	0.25	0.29	0.10	0.50
Korea, Republic of	0.25	0.05	0.27	0.50
Zambia	0.25	0.25	0.33	0.50
Venezuela	0.25	0.16	0.43	0.50
Western Samoa	0.27	0.05	0.30	0.56
Singapore	0.28	0.02	0.29	0.60
Brazil	0.33	0.68	0.21	0.80
Chile	0.33	0.15	0.46	0 80
Argentina	0.39	0.74	0.40	1.00

— Not available.
n.a. Not applicable.

The coefficient of rank correlation between the actual and predicted ri
depreciation of money is 0.25 for the countries in table 11, which indicates th
the combination of variables for turnover and legal central bank independen
can predict a reasonable amount of the cross-country variation in inflation in tl
1980s. Using only the legal variable (and country group dummies), the coefl
cient of rank correlation is merely 0.19. The inclusion of turnover in the coı
struction of an index of central bank independence improves the ability of th
index to predict cross-country variations in inflation. Argentina and Braz
would have been ranked as 55th and 29th from the top in legal independenc
alone, instead of 71st and 69th in the overall index of independence. The larg
change of rankings reflects the effect of the turnover variable, which is very high
for Argentina and Brazil. In most other cases, where including the turnover
variable substantially improved the prediction of a country's inflation, low turn-
over brought the prediction down closer to the actual.

All the industrial countries are above the median of overall independence (D predicted 0.17), and most of the developing countries are below it. Among the countries with overall independence above the median—D estimated below the median—only two (Iceland and Tanzania) have actual D values above 0.17. This indicates that a reasonably high overall independence is highly likely to prevent high inflation. All of the countries with inflation above 50 percent (D 0.33) have less than median central bank independence. For 11 countries, however, the actual D is less than 0.17, even though they have lower-than-median overall central bank independence (estimated D above 0.17). In other words, lower-than-median levels of overall central bank independence do not necessarily lead to high inflation.

These findings are consistent with the view that below-median independence by itself does not necessarily result in high inflation. When there are adverse shocks, however, countries with independence levels within the low range are more likely to develop high and even exceptionally high rates of inflation.

Austria, The Bahamas, Belgium, Luxembourg, Netherlands, and Panama have lower inflation in the 1980s than their central bank independence would indicate, because their monetary policy is dominated by a policy rule fixing their exchange rate to a relatively stable currency. Korea and Japan have lower inflation than their indicators of central bank independence predict, probably because the governments, to which their central banks are subservient, have their own commitment to price stability. These examples demonstrate that high central bank independence is not necessary for price stability.

IV. Conclusions

The notion of central bank independence underlying this study is not unconditional independence from government, but rather the independence to pursue the objective of price stability, even at the cost of other objectives that may be more important to the political authorities. Thus, our measures of independence

include indexes of institutional independence such as appointment procedures, as well as measures of the relative importance attached to price stability in the central bank law and in practice.

Unavoidably, there were subjective or arbitrary decisions in coding, classifying, and weighing legal information and responses to questionnaires. Results of sensitivity analysis offer some reassurance that the main results are robust, but questions about various details may remain. This study aims to contribute to the systematization of future work by committing to a systematic and documented way of characterizing central bank independence.

The study produces four different rankings of independence of central banks. The first is by legal independence, and the second is by governors' turnover rates. The third ranking utilizes responses of specialists to a questionnaire on central bank independence. The fourth ranking is based on an aggregation of the first two.

Legal independence is an important and statistically significant determinant of price stability among industrial countries, but not among developing countries. Within the latter group, some countries have elaborately locked cookie boxes, to borrow a metaphor from the beginning of the article, but some of them either break or undo the lock when they are hungry. An important step in creating an independent central bank, therefore, must involve establishing respect for the central bank charter and management, even when they are not ideal. The rate of turnover of the governors contributes significantly to explaining inflation, and it is even more important in explaining variations in inflation across the overall sample of countries. An inflation-based index of overall central bank independence—combining legal and turnover information—contributes significantly to explaining cross-country variations in the rate of inflation. These results seem robust to possible biases due to reverse causality. The results imply that the discrepancies between actual and legal independence are wider on average in developing countries than in the industrial countries. The turnover rate was not significant in explaining variations of inflation within the industrial group.

Preliminary evidence from the 1980s suggests that when questionnaire variables are used to explain variations in inflation, there is some additional information in the rate of governors' turnover but not in the legal variables.

There seems to be a vicious circle between inflation and the lack of central bank independence, which deserves fuller investigation. Preliminary results here indicate a two-way (positive in both directions) causality between inflation and turnover of central bank governors, a proxy for lack of independence. Lower independence induces higher future inflation, which, in turn, reduces the subsequent actual level of central bank independence, and so on. Success in controlling inflation, however, seems to enhance the independence of central banks.

Finally, central bank independence is only one of several institutional devices for ensuring price stability. Some of the structural and political determinants of central bank independence are discussed in Cukierman (1992a, chap. 23, and 1992b).

Table A-1. *Disaggregated and Aggregated Legal Central Bank Independence Variables, by Country and Decade*

Economy and decade	*Disaggregated legal variables* — *CEO variables*: Term of office (1a)	Who appoints (1b)	Dis-missal (1c)	Other offices (1d)	*Policy formulation variables*: Who formulates (2a)	Final authority (2b)	Role in budget (2c)	Central bank objectives variable (3)	*Limitations on lending variables*: Advances (4a)	Securitized lending (4b)	Terms of lending (4c)	Potential borrowers (4d)	Type of limit (4e)	Maturity of loans (4f)	Interest rates (4g)	Primary market (4h)	Aggregate legal central bank independence variable
Argentina																	
1972–79	0.25	0.25	0.83	1.00	0.33	0.00	0.00	0.40	0.33	0.33	0.33	1.00	0.33	1.00	0.50	0.00	0.40
1980–89	0.25	0.25	0.83	1.00	0.33	0.00	0.00	0.40	0.33	0.33	0.33	1.00	0.33	1.00	0.50	0.00	0.40
Australia																	
1960–71	0.75	0.00	0.83	1.00	0.33	0.20	0.00	0.40	0.33	0.00	0.33	0.00	—	0.67	1.00	0.00	0.36
1972–79	0.75	0.00	0.83	1.00	0.33	0.20	0.00	0.40	0.33	0.00	0.33	0.00	—	0.67	1.00	0.00	0.36
1980–89	0.75	0.00	0.83	1.00	0.33	0.20	0.00	0.40	0.33	0.00	0.33	0.00	—	0.67	1.00	0.00	0.36
Austria																	
1950–59	0.50	0.00	0.83	1.00	0.67	0.60	0.00	0.60	1.00	0.67	0.33	1.00	1.00	1.00	1.00	0.00	0.65
1960–71	0.50	0.00	0.83	1.00	0.67	0.60	0.00	0.60	1.00	0.67	0.33	1.00	1.00	1.00	1.00	0.00	0.65
1972–79	0.50	0.00	0.83	1.00	1.00	0.60	0.00	0.60	1.00	0.67	0.33	0.33	0.33	1.00	1.00	0.00	0.61
1980–89	0.50	0.00	0.83	1.00	1.00	0.60	0.00	0.60	1.00	0.67	0.33	0.33	0.33	1.00	1.00	0.00	0.61
Bahamas, The																	
1972–79	0.50	0.00	0.83	0.50	0.33	—	0.00	0.60	0.33	0.33	0.33	1.00	0.33	1.00	0.25	0.00	0.41
1980–89	0.50	0.00	0.83	0.50	0.33	—	0.00	0.60	0.33	0.33	0.33	1.00	0.33	1.00	0.25	0.00	0.41
Barbados																	
1972–79	0.50	0.00	0.83	0.00	0.33	0.20	0.00	0.80	0.33	0.33	0.33	0.67	—	0.33	0.25	0.00	0.38
1980–89	0.50	0.00	0.83	0.00	0.33	0.20	0.00	0.80	0.33	0.33	0.33	0.67	—	0.33	0.25	0.00	0.38
Belgium																	
1950–59	0.50	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.33	—	1.00	0.50	0.00	0.15
1960–71	0.50	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.00	0.00	0.33	0.33	—	1.00	0.50	0.00	0.15
1972–79	0.50	0.00	0.00	0.50	0.00	0.20	0.00	0.00	0.00	0.00	0.33	0.33	—	1.00	0.50	0.00	0.17
1980–89	0.50	0.00	0.00	0.50	0.00	0.20	0.00	0.00	0.00	0.00	0.33	0.33	—	1.00	0.50	0.00	0.17
Bolivia																	
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

[illegible]	0.00	0.50	0.00	0.00	0.33	—	0.00	0.00	0.67	0.00	0.00	1.00	—	0.00	0.25	0.00	0.21
1972–79	0.00	0.50	0.00	0.00	0.33	—	0.00	0.00	0.67	0.00	0.00	1.00	—	0.00	0.25	0.00	0.21
1980–89	0.00	0.50	0.00	0.00	0.33	—	0.00	0.00	0.67	0.00	0.00	1.00	—	0.00	0.25	0.00	0.21
Canada																	
1950–59	0.75	0.75	0.83	1.00	0.33	0.20	0.00	0.20	0.33	0.67	0.44	0.67	0.33	0.67	0.75	0.00	0.45
1960–71	0.75	0.75	0.83	1.00	0.33	0.20	0.00	0.20	0.33	0.33	0.67	0.67	0.33	0.67	0.75	0.00	0.45
1972–79	0.75	0.75	0.83	1.00	0.33	0.20	0.00	0.20	0.33	0.33	0.67	0.67	0.33	0.67	0.75	0.00	0.45
1980–89	0.75	0.75	0.83	1.00	0.33	0.20	0.00	0.20	0.33	0.33	0.67	0.67	0.33	0.67	0.75	0.00	0.45
Chile																	
1950–59	0.00	1.00	1.00	0.50	0.67	0.00	0.00	0.20	0.00	0.00	0.67	0.00	—	0.00	0.25	0.00	0.26
1960–71	0.00	1.00	1.00	0.50	0.67	0.00	0.00	0.20	0.00	0.00	0.67	0.00	—	0.00	0.25	0.00	0.26
1972–79	0.50	0.00	0.83	0.50	0.67	0.20	0.00	0.80	0.33	0.33	0.67	1.00	—	0.00	0.25	0.00	0.46
1980–89	0.50	0.00	0.83	0.50	0.67	0.20	0.00	0.80	0.33	0.33	0.67	1.00	—	0.00	0.25	0.00	0.46
China																	
1950–59	0.50	—	—	0.00	—	0.80	0.00	0.20	—	0.33	0.00	—	—	0.00	0.25	1.00	0.29
1960–71	0.50	—	—	0.00	—	0.80	0.00	0.20	—	0.33	0.00	—	—	0.00	0.25	1.00	0.29
1972–79	0.50	—	—	0.00	—	0.80	0.00	0.20	—	0.33	0.00	—	—	0.00	0.25	1.00	0.29
1980–89	0.50	—	—	0.00	—	0.80	0.00	0.20	—	0.33	0.00	—	—	0.00	0.25	1.00	0.29
Colombia																	
1960–71	0.00	0.75	0.83	0.00	0.00	0.20	0.00	0.00	0.67	0.00	0.33	0.00	0.67	0.67	0.25	0.00	0.27
1972–79	0.00	0.75	0.83	0.00	0.00	0.20	0.00	0.00	0.67	0.00	0.33	0.00	0.67	0.67	0.25	0.00	0.27
1980–89	0.00	0.75	0.83	0.00	0.00	0.20	0.00	0.00	0.67	0.00	0.33	0.00	0.67	0.67	0.25	0.00	0.27
Costa Rica																	
1960–71	—	1.00	0.67	1.00	—	—	0.00	0.60	0.67	0.33	0.33	0.33	0.00	0.67	0.25	0.00	0.47
1972–79	—	1.00	0.67	1.00	—	—	0.00	0.60	0.67	0.33	0.33	0.33	0.00	0.67	0.25	0.00	0.47
1980–89	—	1.00	0.67	1.00	—	—	0.00	0.60	0.67	0.33	0.33	0.33	0.00	0.67	0.25	0.00	0.47
Denmark																	
1950–59	0.00	0.00	0.33	0.00	—	1.00	0.00	0.60	1.00	0.33	0.67	0.00	—	1.00	0.25	0.00	0.50
1960–71	0.00	0.00	0.33	0.00	—	1.00	0.00	0.60	1.00	0.33	0.67	0.00	—	1.00	0.25	0.00	0.50
1972–79	0.00	0.00	0.33	0.00	—	1.00	0.00	0.60	1.00	0.33	0.67	0.00	—	1.00	0.25	0.00	0.50
1980–89	0.00	0.00	0.33	0.00	—	1.00	0.00	0.60	1.00	0.33	0.67	0.00	—	1.00	0.25	0.00	0.50

(Table continues on the following page.)

Table A-1. *(continued)*

Economy and decade	CEO variables: Term of office (1a)	Who appoints (1b)	Dismissal (1c)	Other offices (1d)	Policy formulation variables: Who formulates (2a)	Final authority (2b)	Role in budget (2c)	Central bank objectives variable (3)	Limitations on lending variables: Advances (4a)	Securitized lending (4b)	Terms of lending (4c)	Potential borrowers (4d)	Type of limit (4e)	Maturity of loans (4f)	Interest rates (4g)	Primary market (4h)	Aggregate legal central bank independence variable
	Disaggregated legal variables																
Egypt																	
1950–59	0.50	0.50	1.00	1.00	0.33	0.00	0.00	0.60	0.67	0.67	0.33	1.00	0.33	0.00	0.25	0.00	0.52
1960–71	0.50	0.50	1.00	1.00	0.33	0.00	0.00	0.60	0.67	0.67	0.33	1.00	0.33	0.00	0.25	0.00	0.52
1972–79	0.50	0.50	1.00	0.00	0.33	0.00	0.00	0.60	0.67	0.67	0.33	1.00	0.67	0.67	0.25	0.00	0.49
1980–89	0.50	0.50	1.00	0.00	0.33	0.00	0.00	0.60	0.67	0.67	0.33	1.00	0.67	0.67	0.25	0.00	0.49
Ethiopia																	
1960–71	0.50	0.25	0.00	0.00	—	—	—	0.60	—	—	—	—	—	0.00	0.25	0.00	—
1972–79	0.00	0.25	0.00	0.00	—	1.00	1.00	0.00	0.33	0.33	0.67	1.00	0.33	0.67	0.50	0.00	0.40
1980–89	0.00	0.25	0.00	0.00	—	1.00	1.00	0.00	0.33	0.33	0.67	1.00	0.33	0.67	0.50	0.00	0.40
Finland																	
1950–59	1.00	0.00	—	0.00	—	0.00	0.00	0.80	0.00	0.00	0.67	—	—	0.00	0.25	0.00	0.28
1960–71	1.00	0.00	—	0.00	—	0.00	0.00	0.80	0.00	0.00	0.67	—	—	0.00	0.25	0.00	0.28
1972–79	1.00	0.00	—	0.00	—	0.00	0.00	0.80	0.00	0.00	0.67	—	—	0.00	0.25	0.00	0.28
1980–89	1.00	0.00	—	0.00	—	0.00	0.00	0.80	0.00	0.00	0.67	—	—	0.00	0.25	0.00	0.28
France																	
1950–59	0.00	0.25	1.00	0.50	0.33	0.60	0.00	0.20	0.33	0.00	0.33	0.00	1.00	0.00	0.00	0.00	0.28
1960–71	0.00	0.25	1.00	0.50	0.33	0.60	0.00	0.20	0.33	0.00	0.33	1.00	1.00	1.00	0.00	0.00	0.36
1972–79	0.00	0.25	1.00	0.50	0.67	0.60	0.00	0.00	0.00	0.00	0.33	1.00	—	—	—	0.00	0.24
1980–89	0.00	0.25	1.00	0.50	0.67	0.60	0.00	0.00	0.00	0.00	0.33	1.00	—	—	—	0.00	0.24
Germany, Fed. Rep. of																	
1950–59	1.00	0.75	1.00	0.00	0.67	1.00	0.00	1.00	0.67	0.67	0.67	0.33	1.00	1.00	0.25	0.00	0.69
1960–71	1.00	0.75	1.00	0.00	0.67	1.00	0.00	1.00	0.67	0.67	0.67	0.33	1.00	1.00	0.25	0.00	0.69
1972–79	1.00	0.75	1.00	0.00	0.67	1.00	0.00	1.00	0.67	0.67	0.67	0.33	1.00	1.00	0.25	0.00	0.69
1980–89	1.00	0.75	1.00	0.00	0.67	1.00	0.00	1.00	0.67	0.67	0.67	0.33	1.00	1.00	0.25	0.00	0.69

1950–59	0.50	0.00	0.83	0.50	0.33	0.00	0.00	0.60	0.67	0.67	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
1960–71	0.50	0.00	0.83	0.50	0.33	0.00	0.00	0.60	0.00	0.33	0.33	0.00	0.33	1.00	0.50	0.00	0.31
1972–79	0.50	0.00	0.83	0.50	0.33	0.00	0.00	0.60	0.00	0.33	0.33	0.00	0.33	1.00	0.50	0.00	0.31
1980–89	0.50	0.00	0.83	0.50	0.33	0.00	0.00	0.60	0.00	0.33	0.33	0.00	0.33	1.00	0.50	0.00	0.31
Greece																	
1950–59	0.25	0.75	0.67	0.50	0.33	0.60	0.00	0.80	0.67	0.67	0.33	0.33	1.00	0.67	0.75	0.00	0.56
1960–71	0.25	0.75	0.67	0.50	0.33	0.60	0.00	0.80	0.33	0.67	0.33	0.33	1.00	0.67	0.75	0.00	0.51
1972–79	0.25	0.75	0.67	0.50	0.33	0.60	0.00	0.80	0.67	0.67	0.33	0.00	1.00	1.00	0.75	0.00	0.55
1980–89	0.25	0.75	0.67	0.50	0.33	0.60	0.00	0.80	0.67	0.67	0.33	0.00	1.00	1.00	0.75	0.00	0.55
Honduras																	
1950–59	0.75	0.00	0.83	1.00	0.33	1.00	0.00	0.00	—	0.33	0.67	0.33	0.33	0.67	0.25	0.00	0.43
1960–71	0.75	0.00	0.83	1.00	0.33	1.00	0.00	0.00	—	0.33	0.67	0.33	0.33	0.67	0.25	0.00	0.43
1972–79	0.75	0.00	0.83	1.00	0.33	1.00	0.00	0.00	—	0.33	0.67	0.33	0.33	0.67	0.25	0.00	0.43
1980–89	0.75	0.00	0.83	1.00	0.33	1.00	0.00	0.00	—	0.33	0.67	0.33	0.33	0.67	0.25	0.00	0.43
Hungary																	
1950–59	0.50	0.25	1.00	0.00	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.33	1.00	0.67	1.00	0.00	0.24
1960–71	0.50	0.25	1.00	0.00	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.33	1.00	0.67	1.00	0.00	0.24
1972–79	0.50	0.25	1.00	0.00	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.33	1.00	0.67	1.00	0.00	0.24
1980–89	0.50	0.25	1.00	0.00	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.33	1.00	0.67	1.00	0.00	0.24
Iceland																	
1950–59	1.00	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
1960–71	1.00	0.75	0.83	0.50	0.33	0.20	0.00	0.40	0.00	0.00	0.33	1.00	—	0.33	0.25	0.00	0.34
1972–79	1.00	0.75	0.83	0.50	0.33	0.20	0.00	0.40	0.00	0.00	0.33	1.00	—	0.33	0.25	0.00	0.34
1980–89	1.00	0.75	0.83	0.50	0.33	0.20	0.00	0.40	0.00	0.00	0.33	1.00	—	0.33	0.25	0.00	0.34
India																	
1950–59	0.50	0.25	0.83	0.50	—	0.00	0.00	0.40	0.00	0.00	0.33	0.33	1.00	0.33	0.25	0.00	0.25
1960–71	0.50	0.25	0.83	0.50	—	0.00	0.00	0.40	0.33	0.00	0.67	0.33	1.00	0.33	0.25	0.00	0.34
1972–79	0.50	0.25	0.83	0.50	—	0.00	0.00	0.40	0.33	0.00	0.67	0.33	1.00	0.33	0.25	0.00	0.34
1980–89	0.50	0.25	0.83	0.50	—	0.00	0.00	0.40	0.33	0.00	0.67	0.33	1.00	0.33	0.25	0.00	0.34
Indonesia																	
1950–59	0.50	0.25	0.00	0.50	0.67	0.00	0.00	0.00	0.33	0.33	0.00	1.00	0.33	0.00	0.25	0.00	0.24
1960–71	0.50	0.25	0.00	0.50	0.67	0.00	0.00	0.40	0.33	0.33	0.00	1.00	0.33	0.00	0.25	0.00	0.30
1972–79	0.50	0.25	0.00	0.50	0.67	0.20	0.00	0.40	0.00	0.00	0.33	1.00	0.33	0.33	0.25	0.00	0.27
1980–89	0.50	0.25	0.00	0.50	0.67	0.20	0.00	0.40	0.00	0.00	0.33	1.00	0.33	0.33	0.25	0.00	0.27

(*Table continues on the following page.*)

Table A-1. *(continued)*

Economy and decade	Disaggregated legal variables																Aggregate legal central bank independence variable
	CEO variables				Policy formulation variables			Central bank objectives variable	Limitations on lending variables								
	Term of office (1a)	Who appoints (1b)	Dis-missal (1c)	Other offices (1d)	Who formu-lates (2a)	Final authority (2b)	Role in budget (2c)	(3)	Advances (4a)	Securitized lending (4b)	Terms of lending (4c)	Potential borrowers (4d)	Type of limit (4e)	Maturity of loans (4f)	Interest rates (4g)	Primary market (4h)	
Ireland																	
1950–59	0.75	0.50	0.83	1.00	—	—	0.00	0.80	—	0.00	0.33	0.33	—	0.67	0.75	0.00	0.44
1960–71	0.75	0.50	0.83	1.00	—	—	0.00	0.80	—	0.00	0.33	0.33	—	0.67	0.75	0.00	0.44
1972–79	0.75	0.50	0.83	1.00	—	—	0.00	0.80	—	0.00	0.33	0.33	—	0.67	0.75	0.00	0.44
1980–89	0.75	0.50	0.83	1.00	—	—	0.00	0.80	—	0.00	0.33	0.33	—	0.67	0.75	0.00	0.44
Israel																	
1950–59	0.50	0.50	0.50	0.50	0.67	0.20	0.00	0.40	0.33	0.00	0.67	1.00	0.00	0.67	0.25	0.00	0.39
1960–71	0.50	0.50	0.50	0.50	0.67	0.20	0.00	0.40	0.33	0.00	0.67	1.00	0.00	0.67	0.25	0.00	0.39
1972–79	0.50	0.50	0.50	0.50	0.67	0.20	0.00	0.40	0.33	0.00	0.67	1.00	0.00	0.67	0.25	0.00	0.39
1980–89	0.50	0.50	0.50	0.50	0.67	0.20	0.00	0.40	0.33	0.00	0.67	1.00	0.00	0.67	0.25	0.00	0.39
Italy																	
1950–59	0.00	0.75	0.67	1.00	—	—	0.00	0.20	0.33	0.00	0.33	—	0.00	0.00	0.25	0.00	0.25
1960–71	0.00	0.75	0.67	1.00	—	—	0.00	0.20	0.33	0.00	0.33	—	0.00	0.00	0.25	0.00	0.25
1972–79	0.00	0.75	0.67	1.00	—	—	0.00	0.20	0.33	0.00	0.33	—	0.00	0.00	0.25	0.00	0.25
1980–89	0.00	0.75	0.67	1.00	—	—	0.00	0.20	0.33	0.00	0.33	—	0.00	0.00	0.25	0.00	0.25
Japan																	
1950–59	0.50	0.25	0.83	0.50	0.67	0.00	0.00	0.00	0.00	0.00	0.33	—	—	0.00	0.25	0.00	0.18
1960–71	0.50	0.25	0.83	0.50	0.67	0.00	0.00	0.00	0.00	0.00	0.33	—	—	0.00	0.25	0.00	0.18
1972–79	0.50	0.25	0.83	0.50	0.67	0.00	0.00	0.00	0.00	0.00	0.33	—	—	0.00	0.25	0.00	0.18
1980–89	0.50	0.25	0.83	0.50	0.67	0.00	0.00	0.00	0.00	0.00	0.33	—	—	0.00	0.25	0.00	0.18
Kenya																	
1960–71	0.25	0.00	0.83	0.50	—	0.20	0.00	0.40	0.67	0.67	0.33	0.33	1.00	0.67	0.75	0.00	0.44
1972–79	0.25	0.00	0.83	0.50	—	0.20	0.00	0.40	0.67	0.67	0.33	0.33	1.00	0.67	0.75	0.00	0.44
1980–89	0.25	0.00	0.83	0.50	—	0.20	0.00	0.40	0.67	0.67	0.33	0.33	1.00	0.67	0.75	0.00	0.44

1950–59	0.25	0.25	0.83	0.00	0.33	—	0.00	0.60	0.33	0.00	0.33	0.00	—	0.67	0.25	0.00	0.30
1960–71	0.25	0.25	0.83	0.00	0.33	—	0.00	0.60	0.33	0.00	0.33	0.00	—	0.67	0.25	0.00	0.30
1972–79	0.25	0.25	0.83	0.50	0.33	—	0.00	0.60	0.33	0.00	0.33	0.00	—	0.67	0.25	0.00	0.32
1980–89	0.25	0.25	0.83	0.50	0.33	—	0.00	0.60	0.00	0.00	0.33	0.00	—	0.67	0.25	0.00	0.27
Lebanon																	
1950–59	0.75	0.25	0.83	1.00	0.67	—	0.00	0.00	0.67	0.00	0.33	0.33	0.33	1.00	1.00	0.00	0.40
1960–71	0.75	0.25	0.83	1.00	0.67	—	0.00	0.00	0.67	0.00	0.33	0.33	0.33	1.00	1.00	0.00	0.40
1972–79	0.75	0.25	0.83	1.00	0.67	—	0.00	0.00	0.67	0.00	0.33	0.33	0.33	1.00	1.00	0.00	0.40
1980–89	0.75	0.25	0.83	1.00	0.67	—	0.00	0.00	0.67	0.00	0.33	0.33	0.33	1.00	1.00	0.00	0.40
Luxembourg																	
1980–89	0.75	0.25	0.83	0.00	0.33	—	0.00	0.60	0.00	0.00	0.33	1.00	—	1.00	0.25	0.00	0.33
Malaysia																	
1960–71	0.50	0.00	0.83	0.00	0.00	0.20	0.00	0.60	0.33	—	0.67	0.00	0.33	0.67	0.25	0.00	0.36
1972–79	0.50	0.00	0.83	0.00	0.00	0.20	0.00	0.60	0.33	—	0.67	0.00	0.33	0.67	0.25	0.00	0.36
1980–89	0.50	0.00	0.83	0.00	0.00	0.20	0.00	0.60	0.33	—	0.67	0.00	0.33	0.67	0.25	0.00	0.36
Malta																	
1960–71	0.50	0.50	0.83	1.00	0.00	0.20	0.00	0.40	0.67	0.00	0.33	1.00	0.33	1.00	0.25	0.00	0.44
1972–79	0.50	0.50	0.83	1.00	0.00	0.20	0.00	0.40	0.67	0.00	0.33	1.00	0.33	1.00	0.25	0.00	0.44
1980–89	0.50	0.50	0.83	1.00	0.00	0.20	0.00	0.40	0.67	0.00	0.33	1.00	0.33	1.00	0.25	0.00	0.44
Mexico																	
1950–59	—	0.00	0.83	1.00	0.00	0.20	0.00	0.00	0.00	0.00	0.33	1.00	—	0.67	0.25	0.00	0.25
1960–71	—	1.00	0.83	1.00	0.67	0.20	0.00	0.00	0.00	0.00	0.33	1.00	—	0.67	0.25	0.00	0.34
1972–79	—	1.00	0.83	1.00	0.67	0.20	0.00	0.00	0.00	0.00	0.33	1.00	—	0.67	0.25	0.00	0.34
1980–89	—	1.00	0.83	1.00	0.67	0.20	0.00	0.00	0.00	0.00	0.33	1.00	—	0.67	0.25	0.00	0.34
Morocco																	
1960–71	—	0.25	0.00	0.00	—	—	0.00	0.20	0.33	0.00	0.00	0.33	0.33	0.67	0.25	0.00	0.14
1972–79	—	0.25	0.00	0.00	—	—	0.00	0.20	0.33	0.00	0.00	0.33	0.33	0.67	0.25	0.00	0.14
1980–89	—	0.25	0.00	0.00	—	—	0.00	0.20	0.33	0.00	0.00	0.33	0.33	0.67	0.25	0.00	0.14
Nepal																	
1950–59	0.50	0.25	0.00	0.00	—	0.00	0.00	0.20	0.00	0.00	0.33	1.00	—	0.67	0.25	0.00	0.18
1960–71	0.50	0.25	0.00	0.00	—	0.00	0.00	0.20	0.00	0.00	0.33	1.00	—	0.67	0.25	0.00	0.18
1972–79	0.50	0.25	0.00	0.00	—	0.00	0.00	0.20	0.00	0.00	0.33	1.00	—	0.67	0.25	0.00	0.18
1980–89	0.50	0.25	0.00	0.00	—	0.00	0.00	0.20	0.00	0.00	0.33	1.00	—	0.67	0.25	0.00	0.18

(Table continues on the following page.)

Table A-1. *(continued)*

Economy and decade	Term of office (1a)	Who appoints (1b)	Dis-missal (1c)	Other offices (1d)	Who formu-lates (2a)	Final authority (2b)	Role in budget (2c)	Central bank objectives variable (3)	Advances (4a)	Securitized lending (4b)	Terms of lending (4c)	Potential borrowers (4d)	Type of limit (4e)	Maturity of loans (4f)	Interest rates (4g)	Primary market (4h)	Aggregate legal central bank independence variable
	Disaggregated legal variables																
	CEO variables				*Policy formulation variables*				*Limitations on lending variables*								
Netherlands																	
1950–59	0.75	0.00	0.17	1.00	0.33	0.20	0.00	0.80	0.67	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.42
1960–71	0.75	0.00	0.17	1.00	0.33	0.20	0.00	0.80	0.67	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.42
1972–79	0.75	0.00	0.17	1.00	0.33	0.20	0.00	0.80	0.67	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.42
1980–89	0.75	0.00	0.17	1.00	0.33	0.20	0.00	0.80	0.67	0.00	0.00	1.00	1.00	0.00	0.00	0.00	0.42
New Zealand																	
1950–59	0.50	0.00	0.83	0.00	0.00	0.00	0.00	0.00	—	—	0.00	0.33	0.33	1.00	0.75	0.00	0.18
1960–71	0.50	0.00	0.83	1.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	1.00	—	0.00	0.50	0.00	0.24
1972–79	0.50	0.00	0.83	1.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	1.00	—	0.00	0.50	0.00	0.24
1980–89	0.50	0.00	0.83	1.00	0.00	0.00	0.00	0.40	0.00	0.00	0.00	1.00	—	0.00	0.50	0.00	0.24
Nicaragua																	
1960–71	0.00	0.00	0.83	1.00	1.00	—	0.00	0.00	1.00	0.33	0.67	0.00	0.33	0.67	0.50	0.00	0.45
1972–79	0.00	0.00	0.83	1.00	1.00	—	0.00	0.00	1.00	0.33	0.67	0.00	0.33	0.67	0.50	0.00	0.45
1980–89	0.00	0.00	0.83	1.00	1.00	—	0.00	0.00	1.00	0.33	0.67	0.00	0.33	0.67	0.50	0.00	0.45
Nigeria																	
1960–71	0.50	0.00	0.83	0.50	0.33	0.20	0.00	0.60	0.33	0.33	0.33	0.00	0.33	0.67	0.75	0.00	0.37
1972–79	0.50	0.00	0.83	0.50	0.33	0.20	0.00	0.60	0.33	0.33	0.33	0.00	0.33	0.67	0.75	0.00	0.37
1980–89	0.50	0.00	0.83	0.50	0.33	0.20	0.00	0.60	0.33	0.33	0.33	0.00	0.33	0.67	0.75	0.00	0.37
Norway																	
1950–59	0.75	0.00	0.33	1.00	0.00	0.20	0.00	0.00	—	—	0.00	0.00	—	0.67	0.25	0.00	0.20
1960–71	0.75	0.00	0.33	1.00	0.00	0.20	0.00	0.00	0.00	0.00	0.00	0.00	—	0.67	0.25	0.00	0.15
1972–79	0.75	0.00	0.33	1.00	0.33	0.20	0.00	0.00	0.00	0.00	0.00	0.00	—	0.67	0.50	0.00	0.17
1980–89	0.75	0.00	0.33	1.00	0.33	0.20	0.00	0.00	0.00	0.00	0.00	0.00	—	0.67	0.50	0.00	0.17
Pakistan																	
1950–59	0.50	0.25	0.83	0.50	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.33	—	0.67	0.25	0.00	0.21
1960–71	0.50	0.25	0.83	0.50	0.00	0.00	0.00	0.40	0.00	0.00	0.00	0.33	—	0.67	0.25	0.00	0.21
1972–79	0.50	0.25	0.83	0.50	0.00	0.00	0.00	0.40	0.00	0.00	0.00	[illegible]	—	[illegible]	[illegible]	[illegible]	[illegible]
1980–89	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Panama																	
1950–59	—	0.25	0.83	1.00	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.33	—	0.00	0.25	0.00	0.24
1960–71	—	0.25	0.83	1.00	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.33	—	0.00	0.25	0.00	0.24
1972–79	—	0.25	0.83	1.00	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.00	—	0.00	0.25	0.00	0.22
1980–89	—	0.25	0.83	1.00	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.00	—	0.00	0.25	0.00	0.22
Peru																	
1960–71	0.00	1.00	0.83	1.00	0.67	0.20	0.00	0.40	0.00	0.67	0.67	0.33	0.33	1.00	0.25	0.00	0.43
1972–79	0.00	1.00	0.83	1.00	0.67	0.20	0.00	0.40	0.00	0.67	0.67	0.33	0.33	1.00	0.25	0.00	0.43
1980–89	0.00	1.00	0.83	1.00	0.67	0.20	0.00	0.40	0.00	0.67	0.67	0.33	0.33	1.00	0.25	0.00	0.43
Philippines																	
1950–59	0.75	0.00	0.83	0.50	0.67	0.20	0.00	1.00	0.33	0.00	0.33	0.67	0.33	0.33	0.25	0.00	0.43
1960–71	0.75	0.00	0.83	0.50	0.67	0.20	0.00	1.00	0.33	0.00	0.33	0.67	0.33	0.33	0.25	0.00	0.43
1972–79	0.75	0.00	0.83	0.50	0.67	0.20	0.00	1.00	0.33	0.00	0.33	0.67	0.33	0.33	0.25	0.00	0.43
1980–89	0.75	0.00	0.83	0.50	0.67	0.20	0.00	1.00	0.33	0.00	0.33	0.67	0.33	0.33	0.25	0.00	0.43
Poland																	
1950–59	0.50	0.25	0.00	0.50	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.00	—	0.67	0.25	0.00	0.14
1960–71	0.50	0.25	0.00	0.50	0.33	0.00	0.00	0.00	0.00	0.00	0.33	0.00	—	0.67	0.25	0.00	0.14
1972–79	0.50	0.25	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.33	0.00	—	0.00	0.25	0.00	0.10
1980–89	0.50	0.25	0.00	0.00	0.67	0.00	0.00	0.00	0.00	0.00	0.33	0.00	—	0.00	0.25	0.00	0.10
Portugal																	
1972–79	0.50	0.25	1.00	1.00	0.33	0.40	0.00	0.60	0.33	0.33	0.67	0.67	0.00	0.67	0.25	0.00	0.48
1980–89	0.00	0.25	1.00	0.00	0.33	0.60	0.00	0.00	0.67	0.67	0.67	0.00	1.00	1.00	0.25	0.00	0.41
Qatar																	
1972–79	0.50	0.25	0.83	0.50	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.33	—	0.00	0.25	0.00	0.20
1980–89	0.50	0.25	0.83	0.50	0.33	0.00	0.00	0.40	0.00	0.00	0.00	0.33	—	0.00	0.25	0.00	0.20
Romania																	
1950–59	0.50	0.25	1.00	1.00	0.67	0.60	0.00	0.60	0.67	0.33	0.67	0.00	0.67	0.67	0.00	0.00	0.53
1960–71	0.50	0.25	1.00	1.00	0.67	0.60	0.00	0.60	0.67	0.33	0.67	0.00	0.67	0.67	0.00	0.00	0.53
1972–79	—	0.25	—	0.00	1.00	0.80	0.00	0.20	—	—	0.67	0.00	—	0.00	0.25	0.00	0.30
1980–89	—	0.25	—	0.00	1.00	0.80	0.00	0.20	—	—	0.67	0.00	—	0.00	0.25	0.00	0.30
Singapore																	
1972–79	0.25	0.00	0.83	0.00	—	—	0.00	0.60	—	—	0.33	0.00	—	1.00	0.25	0.00	0.29
1980–89	0.25	0.00	0.83	0.00	—	—	0.00	0.60	—	—	0.33	0.00	—	1.00	0.25	0.00	0.29

(Table continues on the following page.)

Table A-1. *(continued)*

Economy and decade	Term of office (1a)	Who appoints (1b)	Dis-missal (1c)	Other offices (1d)	Who formulates (2a)	Final authority (2b)	Role in budget (2c)	Central bank objectives variable (3)	Advances (4a)	Securitized lending (4b)	Terms of lending (4c)	Potential borrowers (4d)	Type of limit (4e)	Maturity of loans (4f)	Interest rates (4g)	Primary market (4h)	Aggregate legal central bank independence variable
	Disaggregated legal variables																
	CEO variables				*Policy formulation variables*				*Limitations on lending variables*								
South Africa																	
1950–59	0.50	0.00	0.83	0.50	—	—	0.00	0.20	0.00	0.00	0.33	1.00	—	1.00	0.25	0.00	0.25
1960–71	0.50	0.00	0.83	0.50	—	—	0.00	0.20	0.00	0.00	0.33	1.00	—	1.00	0.25	0.00	0.25
1972–79	0.50	0.00	0.83	0.50	—	—	0.00	0.20	0.00	0.00	0.33	1.00	—	1.00	0.25	0.00	0.25
1980–89	0.50	0.00	0.83	0.50	—	—	0.00	0.20	0.00	0.00	0.33	1.00	—	1.00	0.25	0.00	0.25
Spain																	
1950–59	0.00	0.25	0.00	0.50	0.33	0.00	0.00	0.00	0.33	0.00	0.33	0.00	0.00	0.00	0.00	0.00	0.13
1960–71	0.00	0.25	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.09
1972–79	0.00	0.25	0.00	0.00	0.33	0.00	0.00	0.00	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.09
1980–89	0.25	0.00	0.00	1.00	0.33	0.00	0.00	0.60	0.33	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.23
Sweden																	
1950–59	0.00	1.00	—	0.50	—	—	0.00	0.20	0.33	0.00	0.67	0.00	1.00	0.67	0.25	0.00	0.29
1960–71	0.00	1.00	—	0.50	—	—	0.00	0.20	0.33	0.00	0.67	0.00	1.00	0.67	0.25	0.00	0.29
1972–79	0.00	1.00	—	0.50	—	—	0.00	0.20	0.33	0.00	0.67	0.00	1.00	0.67	0.25	0.00	0.29
1980–89	0.00	1.00	—	0.50	—	—	0.00	0.20	0.33	0.00	0.67	0.00	1.00	0.67	0.25	0.00	0.29
Switzerland																	
1950–59	0.75	0.25	—	1.00	—	1.00	0.00	0.00	0.67	0.33	0.67	1.00	—	1.00	0.25	0.00	0.53
1960–71	0.75	0.25	—	1.00	—	1.00	0.00	0.00	0.67	0.33	0.67	1.00	—	1.00	0.25	0.00	0.53
1972–79	0.75	0.25	—	1.00	—	1.00	0.00	0.00	0.67	0.33	0.67	1.00	—	1.00	0.25	0.00	0.53
1980–89	0.75	0.25	—	1.00	—	1.00	0.00	0.00	1.00	—	1.00	1.00	—	1.00	0.25	0.00	0.64
Taiwan																	
1980–89	0.50	0.50	1.00	0.00	0.00	0.00	0.00	0.60	0.00	0.00	0.00	0.33	0.00	0.00	0.00	0.00	0.21
Tanzania																	
1960–71	0.25	0.00	0.83	0.50	0.67	0.20	0.00	0.40	0.67	0.33	0.33	1.00	0.33	0.67	0.75	0.00	0.44
1972–79	0.25	0.00	0.83	0.50	0.67	0.20	0.00	0.40	0.67	0.33	0.33	1.00	0.33	0.67	0.75	0.00	0.44
1980–89	0.25	0.00	0.83	0.50	0.67	0.20	0.00	0.40	0.67	0.33	0.33	1.00	0.33	0.67	0.75	0.00	0.44

1950–59	0.25	0.50	0.00	0.00	0.00	0.20	0.00	0.60	0.33	0.00	0.33	0.33	0.00	0.67	0.25	0.00	0.27
1960–71	0.25	0.50	0.00	0.00	0.00	0.20	0.00	0.60	0.33	0.00	0.33	0.33	0.00	0.67	0.25	0.00	0.27
1972–79	0.25	0.50	0.00	0.00	0.00	0.20	0.00	0.60	0.33	0.00	0.33	0.33	0.00	0.67	0.25	0.00	0.27
1980–89	0.25	0.50	0.00	0.00	0.00	0.20	0.00	0.60	0.33	0.00	0.33	0.33	0.00	0.67	0.25	0.00	0.27
Turkey																	
1950–59	0.50	0.75	1.00	1.00	0.67	0.20	0.00	0.40	0.33	0.00	0.33	0.00	0.67	1.00	0.25	0.00	0.39
1960–71	0.50	0.75	1.00	1.00	0.67	0.20	0.00	0.40	0.33	0.00	0.33	0.00	0.67	1.00	0.25	0.00	0.39
1972–79	0.50	0.75	0.83	0.00	0.33	0.80	0.00	0.60	0.33	0.67	0.33	0.33	0.00	0.67	0.25	0.00	0.46
1980–89	0.50	0.75	0.83	0.00	0.33	0.80	0.00	0.60	0.33	0.67	0.33	0.33	0.00	0.67	0.25	0.00	0.46
Uganda																	
1960–71	0.50	0.50	0.83	0.50	0.00	0.20	0.00	0.40	0.33	—	0.33	0.33	0.33	1.00	0.75	0.00	0.38
1972–79	0.50	0.50	0.83	0.50	0.00	0.20	0.00	0.40	0.33	—	0.33	0.33	0.33	1.00	0.75	0.00	0.38
1980–89	0.50	0.50	0.83	0.50	0.00	0.20	0.00	0.40	0.33	—	0.33	0.33	0.33	1.00	0.75	0.00	0.38
United Kingdom																	
1950–59	0.50	0.00	0.83	1.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	1.00	—	0.00	0.25	0.00	0.21
1960–71	0.50	0.00	0.83	1.00	0.00	0.00	0.00	0.20	0.67	0.67	0.00	1.00	1.00	1.00	0.75	0.00	0.43
1972–79	0.50	0.00	0.83	1.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	1.00	1.00	1.00	0.75	0.00	0.27
1980–89	0.50	0.00	0.83	1.00	0.00	0.00	0.00	0.20	0.00	0.00	0.00	1.00	1.00	1.00	0.75	0.00	0.27
United States																	
1950–59	0.25	0.50	0.00	1.00	—	0.20	0.00	0.40	1.00	0.33	0.33	1.00	—	1.00	0.25	0.00	0.48
1960–71	0.25	0.50	0.00	1.00	—	0.20	0.00	0.40	1.00	0.33	0.33	1.00	—	1.00	0.25	0.00	0.48
1972–79	0.25	0.50	0.00	1.00	—	0.20	0.00	0.40	1.00	0.33	0.33	1.00	—	1.00	0.25	0.00	0.48
1980–89	0.25	0.50	0.00	1.00	—	0.20	0.00	0.40	1.00	0.33	0.33	1.00	—	1.00	0.25	0.00	0.48
Uruguay																	
1950–59	0.25	0.00	—	0.50	0.67	0.00	0.00	0.20	—	—	0.33	0.33	—	0.00	0.25	0.00	0.22
1960–71	0.25	0.00	—	0.50	0.67	0.00	0.00	0.20	—	—	0.33	0.33	—	0.00	0.25	0.00	0.22
1972–79	0.25	0.25	0.83	0.00	0.67	0.20	0.00	0.40	0.00	0.00	0.67	0.00	—	0.00	0.25	0.00	0.24
1980–89	0.25	0.25	0.83	0.00	0.67	0.20	0.00	0.40	0.00	0.00	0.67	0.00	—	0.00	0.25	0.00	0.24
Venezuela																	
1950–59	0.50	0.00	—	0.00	0.00	—	0.00	0.20	—	—	0.67	1.00	—	0.67	0.25	0.00	0.28
1960–71	0.50	0.00	—	0.00	0.00	—	0.00	0.20	—	—	0.67	1.00	—	0.67	0.25	0.00	0.28
1972–79	0.50	0.00	0.83	0.50	0.33	1.00	0.00	0.40	0.67	0.00	0.67	0.00	0.33	0.00	0.50	0.00	0.43
1980–89	0.50	0.00	0.83	0.50	0.33	1.00	0.00	0.40	0.67	0.00	0.67	0.00	0.33	0.00	0.50	0.00	0.43

(*Table continues on the following page.*)

Table A-1. *(continued)*

Economy and decade	CEO variables: Term of office (1a)	CEO variables: Who appoints (1b)	CEO variables: Dis-missal (1c)	CEO variables: Other offices (1d)	Policy formulation variables: Who formu-lates (2a)	Policy formulation variables: Final authority (2b)	Policy formulation variables: Role in budget (2c)	Central bank objectives variable (3)	Limitations on lending variables: Advances (4a)	Limitations on lending variables: Securitized lending (4b)	Limitations on lending variables: Terms of lending (4c)	Limitations on lending variables: Potential borrowers (4d)	Limitations on lending variables: Type of limit (4e)	Limitations on lending variables: Maturity of loans (4f)	Limitations on lending variables: Interest rates (4g)	Limitations on lending variables: Primary market (4h)	Aggregate legal central bank independence variable
Disaggregated legal variables																	
Western Samoa																	
1980–89	0.25	0.25	0.83	0.00	0.00	0.20	0.00	0.40	0.33	0.33	0.33	0.00	0.33	1.00	0.25	0.00	0.30
Yugoslavia																	
1960–71	0.25	0.25	0.83	0.00	0.00	0.20	0.00	0.40	0.00	—	0.00	0.00	—	0.00	0.25	0.00	0.17
1972–79	0.25	0.25	0.83	0.00	0.00	0.20	0.00	0.40	0.00	—	0.00	0.00	—	0.00	0.25	0.00	0.17
1980–89	0.25	0.25	0.83	0.00	0.00	0.20	0.00	0.40	0.00	—	0.00	0.00	—	0.00	0.25	0.00	0.17
Zaire																	
1972–79	0.50	0.00	1.00	0.00	0.67	0.20	0.00	0.60	—	0.00	0.33	0.33	0.33	0.67	0.75	0.00	0.35
1980–89	0.50	0.00	1.00	1.00	0.67	0.60	0.00	0.40	0.33	0.33	0.33	0.33	0.33	0.67	0.75	0.00	0.43
Zambia																	
1960–71	0.50	0.25	0.83	1.00	0.00	0.20	0.00	0.80	0.33	0.33	0.00	0.33	0.33	0.33	0.75	0.00	0.40
1972–79	0.50	0.25	0.83	1.00	0.00	0.20	0.00	0.80	0.33	0.33	0.00	0.33	0.33	0.33	0.75	0.00	0.40
1980–89	0.50	0.25	0.83	0.50	0.33	0.20	0.00	0.40	0.33	0.33	0.00	0.33	0.33	0.33	0.75	0.00	0.33
Zimbabwe																	
1950–59	0.75	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.67	0.33	0.33	0.25	0.00	0.23
1960–71	0.75	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.67	0.33	0.33	0.25	0.00	0.20
1972–79	0.75	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.67	0.33	0.33	0.25	0.00	0.20
1980–89	0.75	0.00	0.00	0.50	0.00	0.00	0.00	0.00	0.33	0.33	0.00	0.67	0.33	0.33	0.25	0.00	0.20

— Not available.

Table A-2. Country Codes for the Questionnaire Variables and the Questionnaire-based Index of Central Bank Independence

Country	Tenure overlap (1)	Limitations on lending (2)	Resolution of conflict (3)	Financial independence: Budget (4a)	Financial independence: Salaries and profits (4b)	Policy targets: Monetary stock (5a)	Policy targets: Interest rate (5b)	Priority of price stability (6)	Function as development bank (7)	Questionnaire-based index of central bank independence
Australia	1.00	0.67	0.00	1.00	0.50	0.33	1.00	1.00	1.00	0.76
Bahamas, The	0.50	0.67	0.50	1.00	1.00	0.00	1.00	0.67	1.00	0.71
Barbados	0.50	0.67	0.00	1.00	0.50	0.00	0.00	1.00	0.67	0.54
Belgium	1.00	0.00	0.00	1.00	1.00	0.00	1.00	0.67	—	0.47
Costa Rica	0.50	0.67	0.50	1.00	1.00	0.67	1.00	1.00	1.00	0.81
Denmark	1.00	1.00	0.50	0.50	1.00	0.00	0.00	0.67	1.00	0.73
Ethiopia	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.33	0.00	0.13
Finland	1.00	1.00	1.00	0.50	0.00	0.00	—	1.00	1.00	0.78
France	0.50	0.33	0.50	0.50	1.00	0.67	1.00	1.00	0.67	0.65
Ireland	0.50	0.33	0.00	0.50	0.00	0.00	1.00	1.00	1.00	0.57
Italy	1.00	0.00	0.50	1.00	1.00	0.67	1.00	1.00	1.00	0.73
Lebanon	0.50	0.33	0.50	1.00	1.00	0.00	1.00	0.33	1.00	0.59
Luxembourg	0.50	0.33	—	1.00	1.00	0.00	1.00	0.67	1.00	0.66
Peru	0.00	0.00	0.00	1.00	0.50	0.33	0.00	0.33	0.33	0.22
South Africa	0.50	0.33	0.50	1.00	0.00	0.67	1.00	—	1.00	0.64
Tanzania	0.50	0.33	0.50	1.00	0.00	0.33	1.00	0.00	0.33	0.38
Turkey	0.50	0.00	0.00	1.00	1.00	0.00	1.00	1.00	0.33	0.44
Uganda	0.50	0.33	0.50	1.00	1.00	0.33	—	1.00	0.33	0.53
Uruguay	0.50	0.33	0.50	1.00	0.50	0.00	—	0.33	1.00	0.49
United Kingdom	0.50	0.33	0.00	0.50	—	0.67	1.00	1.00	1.00	0.64
Germany, Fed. Rep. of	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Yugoslavia	—	0.00	0.00	0.50	0.00	0.33	1.00	—	0.00	0.17
Zaire	0.50	0.33	0.50	1.00	1.00	0.67	1.00	1.00	0.33	0.61

— Not available.

Note: Numbers in parentheses in the column headings correspond to variable numbers in table 4.

References

The word "processed" describes informally reproduced works that may not be commonly available through libraries.

Alesina, Alberto. 1988. "Macroeconomics and Politics." NBER *Macro-economics Annual* 3: 13–52.

Alesina, Alberto, and Guido Tabellini. 1987. "Rules and Discretion with Noncoordinated Monetary and Fiscal Policies." *Economic Inquiry* 25: 619–30.

Aufricht, Hans. 1961. *Central Banking Legislation*. Washington, D.C.: IMF.

———. 1967. *Central Banking Legislation*. Vol. II, *Europe*. Washington, D.C.: IMF.

Bade, Robin, and Michael Parkin. 1980. "Central Bank Laws and Monetary Policy." University of Western Ontario, Department of Economics, London, Ontario. Processed.

Banaian, King, Leroy Laney, and Thomas Willett. 1983. "Central Bank Independence: An International Comparison." *Economic Review: Federal Reserve Bank of Dallas* (March): 1–13.

Bank for International Settlements. 1963. *Eight European Central Banks*. New York: Praeger under the auspices of the BIS.

Bodart, Vincent. 1990. "Central Bank Independence and the Effectiveness of Monetary Policy: A Comparative Analysis." IMF, Central Banking Department, Washington, D.C. Processed.

Bordes, Christian, and M. O. Strauss-Kahn. 1987. "1977–86: Dix ans de Politique D'objectifs en France ou le 'Targeting' a la Francaise." SUERF (Societe Universitaire Europeene de Recherches Financiers) papers on Monetary Policy and Financial Systems 4, Tilburg. Processed.

Bordo, Michael, and Angela Redish. 1987. "Why Did the Bank of Canada Emerge in 1935?" *Journal of Economic History* 47 (June): 405–17.

Cargill, Thomas, and Michael Hutchison. 1990. "Monetary Policy and Political Economy: The Federal Reserve and the Bank of Japan." In Thomas Mayer, ed., *The Political Economy of American Monetary Policy*. New York: Cambridge University Press.

Cukierman, Alex. 1984. *Inflation, Stagflation, Relative Prices, and Imperfect Information*. Cambridge: Cambridge University Press.

———. 1992a. *Central Bank Strategy, Credibility, and Independence: Theory and Evidence*. Cambridge, Mass.: MIT Press.

———. 1992b. "Commitment through Delegation and Its Effect on Policy: The Case of Central Bank Independence." Paper presented at the Conference on Coalitions and Bargaining, Tel Aviv University, Department of Economics. Processed.

Dotsey, Michael. 1986. "Japanese Monetary Policy: A Comparative Analysis." *Economic Review: Federal Reserve Bank of Richmond* 72 (November/December): 12–24.

Effros, Robert, ed. 1982. *Emerging Financial Centers: Legal and Institutional Framework*. Washington, D.C.: IMF.

Eizenga, Weitze. 1983. "The Independence of the Federal Reserve System and of the Netherlands Bank." SUERF (Societe Universitaire Europeene de Recherches Financiers) Series 41A, Tilburg. Processed.

———. 1987. "The Independence of the Deutsche Bundesbank and the Nederlandsche Bank with Regard to Monetary Policy: A Comparative Study." SUERF (Societe Universitaire Europeene de Recherches Financiers) papers on Monetary Policy and Financial Systems 2, Tilburg. Processed.

Epstein, Gerald, and Juliet Schor. 1986. "The Divorce of the Banca d'Italia and the Italian Treasury: A Case Study of Central Bank Independence." Harvard Institute of Economic Research, Discussion Paper Series 1269. Cambridge, Mass. Processed.

Fazio, Antonio. 1991. "Role and Independence of Central Banks." In Patrick Downes and Reza Vaez-Zadeh, eds., *The Evolving Role of Central Banks*. Washington, D.C.: IMF.

Fischer, Stanley. 1986. *Indexing, Inflation, and Economic Policy*. Cambridge, Mass.: MIT Press.

Goodhart, Charles. 1988. *The Evolution of Central Banks*. Cambridge, Mass.: MIT Press.

Grilli, Vittorio, Donato Masciandaro, and Guido Tabellini. 1991. "Political and Monetary Institutions and Public Financial Policies in the Industrial Countries." *Economic Policy* 13: 341–92.

Holtfrerich, Carl-Ludwig. 1988. "Relations between Monetary Authorities and Governmental Institutions: The Case of Germany from the 19th Century to the Present." In Gianni Toniolo, ed., *Central Bank Independence in Historical Perspective*. Berlin: Walter de Gruyter.

Kearney, Colm. 1984. "The British Anti-Inflation Strategy: Implementing Monetarism or Turning to Radcliffe?" SUERF Series 44A. Tilburg. Processed.

Keenan, Peter, and David Mayes. 1987. "Monetary Targets and Monetary Management: The New Zealand Experience." SUERF (Societe Universitaire Europeene de Recherches Financiers) Papers on Monetary Policy and Financial Systems 3, Tilburg. Processed.

Leone, Alfredo. 1991. "Effectiveness and Implications of Limits on Central Bank Credit to the Government." In Patrick Downes and Reza Vaez-Zadeh, eds., *The Evolving Role of Central Banks*. Washington, D.C.: IMF.

Lobel, Arnold. 1972. *Frog and Toad Together*. New York: Harper and Row.

Lohmann, Susanne. 1992. "Optimal Commitment in Monetary Policy: Credibility versus Flexibility." *American Economic Review*. 82 (March): 273–86.

Masciandaro, Donato, and Guido Tabellini. 1988. "Monetary Regimes and Fiscal Deficits: A Comparative Analysis." In H. S. Cheng, ed., *Monetary Policy in Pacific Basin Countries*. Dordrecht: Kluwer Academic Publishers.

Maxfield, Sylvia. 1992. "Private Constituencies and the Power of State Monetary Authorities in Newly Industrializing Countries." Yale University, Department of Political Science, New Haven, Conn. Processed.

Mayer, Thomas, ed. 1990. *The Political Economy of American Monetary Policy*. New York: Cambridge University Press.

Meltzer, Allan. 1991. "U.S. Policy in the Bretton Woods Era." The 1991 Homer Jones Lecture. St. Louis, Mo.

Mittra, Sid. 1977. *Central Bank versus Treasury: An International Study*. Lanham, Md.: University Press of America.

Parkin, Michael. 1986. "Domestic Monetary Institutions and Deficits." In J. M. Buchanan, C. K. Rowley, and R. D. Tollison, eds., *Deficits*. New York: Basil Blackwell.

Rogoff, Kenneth. 1985. "The Optimal Degree of Commitment to an Intermediate Monetary Target." *Quarterly Journal of Economics* 100 (November): 1169–90.

Schokker, E. 1980. "The Central Bank and the State in the Light of European Integration." SUERF (Societe Universitaire Europeene de Recherches Financiers) Series 34A, Tilburg. Processed.

Skanland, Hermod. 1984. *The Central Bank and Political Authorities in Some Industrial Countries*. Oslo: Norges Bank.

Suzuki, Yoshio. 1987. *The Japanese Financial System*. Oxford: Clarendon.

Swinburne, Mark, and Marta Castello-Branco. 1991. "Central Bank Independence and Central Bank Functions." In Patrick Downes and Reza Vaez-Zadeh, eds., *The Evolving Role of Central Banks*. Washington, D.C.: IMF.

Volcker, Paul, Miguel Mancera, and Jean Godeaux. 1991. *Perspectives on the Role of a Central Bank*. Proceedings of a conference held in Beijing, China, January 5–7, 1990. Beijing: Peoples' Bank of China.

Willett, Thomas, ed. 1988. *Political Business Cycles: The Political Economy of Money, Inflation, and Unemployment*. Durham, N.C.: Duke University Press.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 3: 399-422

On the Transmission of World Agricultural Prices

Yair Mundlak and Donald F. Larson

Two questions are asked about the relationship between domestic prices and world prices of agricultural commodities: are variations in world prices transmitted to domestic prices, and do these variations in world prices constitute an important component of variations in domestic prices? Domestic prices are regressed on world prices in various forms, taking into account the possible effects of exchange rates and inflation. The empirical analysis is based on data from the Food and Agriculture Organization of the United Nations for 58 countries for 1968–78 and for the countries of the European Community for 1961–85. The results show that most of the variations in world prices are transmitted and that they constitute the dominant component in the variations of domestic prices.

Agricultural products are on the whole tradable, and every country trades in some agricultural products. In the absence of intervention it is expected that domestic prices of such products will vary with world prices. It is well known, however, that agriculture is subjected to considerable intervention, which creates a gap between world prices and domestic prices and which generates cross-country variations in agricultural prices (see, for example, McCalla 1969; Johnson 1973; Bale and Lutz 1981; Australia 1985; Anderson, Hayami, and Honma 1986; World Bank 1986). This is perhaps why it is sometimes claimed that world prices are irrelevant for the development of agriculture in countries that intervene in the pricing of their agricultural products; see, for instance, the implicit debate in Mellor and Ahmed (1988), Valdés and Siamwalla (1988), and Ahmed (1988). It is therefore natural to ask to what extent such an intervention reduces the influence of world prices on domestic prices.

In this article we examine two major questions. First, what proportion of the variations in world prices is transmitted to domestic prices? Second, what proportion of the variations in domestic prices can be attributed to variations in world prices? Not independently, we also examine empirically the revealed relationships between world prices and the degree of intervention. The outcome of this analysis is crucial for understanding the relationships between domestic and

Yair Mundlak is with the University of Chicago and the International Food Policy Research Institute. Donald F. Larson is with the International Economics Department at the World Bank. The authors are grateful to Ronald C. Duncan and D. Gale Johnson for comments and suggestions on an earlier draft.

world markets, and it has several ramifications. The facts and the ramifications are two distinct subjects, however, and it is important to deal with them separately. This article deals mainly with the empirical analysis.

I. The Framework

The framework draws on the law of one price, where the domestic price of commodity i in year t, P_{it}, is expressed as a product of the world price, P^*_{it}, the nominal exchange rate, E_t, and the tax policy $S_{it} = (1 + \tau_{it})$, where τ is the tax rate (or, if negative, the subsidy). Because countries engage in nontariff market interventions, including quantitative import restrictions, the tax policy term here includes the tariff equivalent of any restriction on domestic prices. The equation for the domestic price is

$$P_{it} = P^*_{it} E_t S_{it}. \tag{1}$$

This formulation ignores differences in product qualities and in transportation, storage, and marketing costs, as well as other domestic nontradable inputs. Also, the equation is based on the assumption that the exchange rate is neither under- nor overvalued so that the difference between domestic and foreign inflation rates is fully reflected in E. To allow for deviations from this assumption and for the effects that are not included in the equation, a disturbance term, denoted by U, is added to the equation.

Equation 1 is rewritten with lowercase letters indicating logs:

$$p_{it} = p^*_{it} + e_t + s_{it} + u_{it} \tag{2}$$

where $u \sim IID(\mu, \sigma^2)$ and $E(eu) = E(su) = E(p^*u) = 0$. The mean of the disturbance, μ, is not necessarily 0, for the reasons given above.

The answer to both questions posed above is obtained, in principle, by computing the following regression:

$$p_{it} = \alpha + \beta p^*_{it} + \gamma e_t + \epsilon_{it}. \tag{3}$$

Equation 2 can be expressed in terms of equation 3, subject to the restrictions H_1: $\beta = 1$, and H_2: $\gamma = 1$. The coefficient β is the elasticity of the domestic price with respect to the world price, to be referred to as the elasticity of transmission. The value of this elasticity is the answer to the first question. A value of 1 implies that the variations in world prices are fully transmitted to the domestic prices, whereas a value of 0 implies no transmission at all. There are several reasons why this elasticity would differ from unity. First, omitted variables, specifically tax-policy variables (s), are correlated with the world price. Second, there may be measurement errors in the world price. Such errors may reflect the fact that the world price used in a given study differs from the one pertinent for the particular country. Third, if the economy is closed, the world price is irrelevant. Of course, very few countries are completely closed, but many countries are

partially closed by means of trade policies, and this may affect the value of the estimate.

The contribution of world prices, measured in domestic currencies, to the variations in domestic prices is given by the value of R^2 of the regression of equation 3. This answers the second question: a low value means that only a small proportion of the variations in domestic prices are accounted for by world prices and exchange rates. The marginal contribution of world prices conditional on the exchange rate is given by the square of the partial correlation coefficient between world and domestic prices.

The foregoing discussion dealt with proportional changes, but it says nothing about differences in levels. This information is contained in the intercept, which need not equal 0. Under the restrictions H_1 and H_2, $\alpha = s_{..} + \mu$, where $s_{..}$ is the sample average, over commodities and time, of s_{it}. Thus the intercept reflects the tax policy and the quantitative importance of the omitted variables from equation 2. More generally, when H_1 or H_2 are not maintained, or if the explanatory variables are measured with error, the intercept will be affected.

II. The Data

The regression was first computed for 58 countries for the period 1968–78, and the sample covered some 60 products. The number of products varied by country. Products not produced in a country were excluded from the analysis. The domestic prices are those given by the Food and Agriculture Organization of the United Nations (FAO) and described by the FAO as follows:

> Farm prices are in theory determined by farmgate or first point-of-sale transaction, when farmers participate in their capacity as sellers of their own products. Of course, data may not always refer to the same selling points, depending on the prevailing institutional setup in the countries. Also, different practices may prevail in regard to individual commodities (FAO 1987, p. 23).

There is a common belief that FAO prices are subject to many problems. This may be the case, but we are unfamiliar with any study that indicates the sources of errors in the FAO data and their quantitative importance, or that substantiates this belief in any other way. One way to have a rough check on the data is to see, as we are doing here, to what extent they are correlated with world prices. If the empirical results showed weak relationships between the domestic and world prices, this could be explained in terms of measurement errors. This is not the case, however, and therefore we think that these data are indeed informative. To double-check, we repeat the analysis in section IV for the European Community, using prices of the U.S. Department of Agriculture reported in Herlihy and others (1989) for 1960–85.

The world price is an export-unit value calculated in nominal U.S. dollars. It

is a ratio of the total world value of exports for each of the commodities divided by the total world exported quantities for the corresponding commodities. We use the framework developed in section I to examine and verify the pertinence of this variable. The exchange rates are annual averages as published by the International Monetary Fund.

III. RESULTS USING FAO DATA

In a cross-commodity comparison, the deviation from unitary transmission elasticity is surprisingly small. The time-series analysis for individual commodities yields somewhat lower values, suggesting that policy does have some smoothing effect. In all cases, the relatively high values of R^2 indicate that world prices constitute a major source of domestic price variations.

Table 1. *Estimated Transmission and Exchange Rate Elasticities, 1968–78*

	Nominal values			*Real values*		
Country	*World prices*	*Exchange rates*	R^2	*World prices*	*Exchange rates*	R^2
Argentina	0.98	0.97	0.94	0.97	0.42	0.71
Australia	0.92	0.63	0.82	0.93	1.04	0.80
Austria	0.98	1.08	0.87	0.98	1.11	0.86
Bangladesh	0.74	0.82	0.66	0.70	0.15	0.50
Belgium and Luxembourg	0.97	1.05	0.79	0.97	1.08	0.79
Brazil	0.87	1.25	0.83	0.86	−1.54	0.69
Burundi	0.87	2.30	0.77	0.88	1.87	0.76
Cameroon	0.89	0.47	0.74	0.88	0.71	0.72
Canada	1.00	−0.20	0.85	1.01	0.35	0.84
Chile	0.94	0.98	0.96	0.95	0.67	0.63
Colombia	0.94	0.85	0.76	0.93	0.60	0.69
Costa Rica	0.93	0.45	0.85	0.92	2.22	0.83
Cyprus	0.92	0.73	0.68	0.93	0.76	0.66
Denmark	1.02	0.81	0.83	1.02	0.86	0.82
Ecuador	1.02	0.07	0.72	0.99	0.53	0.67
Egypt	1.24	2.22	0.74	1.21	−0.04	0.71
El Salvador	0.90	n.a.	0.80	0.89	0.28	0.79
Finland	0.99	3.08	0.89	0.99	1.52	0.88
France	0.95	1.04	0.72	0.95	1.02	0.70
Germany, Fed. Rep. of	0.99	1.23	0.83	0.99	1.29	0.83
Greece	0.90	1.29	0.72	0.90	1.01	0.67
Guatemala	0.90	n.a.	0.82	0.91	1.00	0.80
India	0.77	−0.61	0.71	0.75	0.75	0.66
Ireland	1.02	0.98	0.86	1.03	1.28	0.84
Israel	1.01	0.84	0.78	1.00	0.83	0.67
Italy	0.92	0.68	0.74	0.92	1.47	0.69
Japan	0.89	−0.20	0.79	0.88	0.33	0.76
Kenya	1.07	0.80	0.83	1.08	1.00	0.82
Korea, Rep. of	0.91	1.32	0.74	0.90	−1.03	0.69
Malawi	0.90	−0.86	0.75	n.a.	n.a.	n.a.

Initial Results

Because the variables are in logs, their sample variations represent relative changes. Therefore, the variables have no units, and it is possible to pool the data over all commodities for all years. The estimates of equation 3 appear in the first three columns of table 1. The *t* ratios are all very high (double digits) and therefore are not reported here.

The estimated transmission elasticity (from the world-prices column) varies between 0.74 and 1.24, with a median of 0.952. The values for 49 out of 57 countries fall in the range of 0.85 to 1.07. Thus the discrepancy from 1 is indeed very small. This indicates that the variations in world prices are almost fully transmitted to domestic prices. This is the answer to the first question. Turning to the second question, the values of R^2 vary between 0.66 and 0.96, which

Table 1. *(continued)*

	Nominal values			*Real values*		
Country	*World prices*	*Exchange rates*	R^2	*World prices*	*Exchange rates*	R^2
Malaysia	0.84	0.58	0.77	0.84	0.78	0.76
Mauritius	1.03	1.50	0.90	1.03	0.30	0.88
Mexico	1.02	0.56	0.79	1.02	1.04	0.75
Netherlands	0.98	1.09	0.76	0.98	1.09	0.76
New Zealand	1.03	0.21	0.76	1.04	0.90	0.74
Norway	0.98	1.23	0.89	0.97	1.11	0.88
Pakistan	0.82	0.44	0.76	0.79	−0.02	0.69
Panama	0.93	n.a.	0.77	0.95	0.84	0.75
Peru	0.86	0.98	0.77	0.85	0.58	0.65
Philippines	0.83	0.58	0.75	0.81	0.58	0.69
Portugal	0.96	1.02	0.79	0.97	1.15	0.76
South Africa	0.98	0.37	0.87	0.99	1.51	0.85
Spain	0.93	1.29	0.77	0.93	1.11	0.74
Sri Lanka	0.84	0.67	0.76	0.83	0.58	0.73
Sweden	0.95	2.35	0.82	0.95	1.87	0.82
Switzerland	1.01	0.68	0.74	1.01	0.68	0.73
Syria	0.98	4.79	0.76	0.97	0.86	0.72
Tanzania	0.97	1.41	0.81	0.98	1.16	0.79
Thailand	0.89	−0.86	0.80	0.89	0.75	0.78
Trinidad	1.01	1.04	0.66	1.01	1.05	0.63
Turkey	0.96	0.94	0.76	0.93	0.10	0.69
United Kingdom	0.96	0.75	0.89	0.96	1.37	0.88
United States	1.01	n.a.	0.82	1.02	0.00	0.81
Uruguay	0.81	0.98	0.94	0.79	0.97	0.72
Venezuela	0.94	5.13	0.71	0.92	−2.02	0.69
Yugoslavia	1.01	1.00	0.83	1.01	1.06	0.79
Zambia	0.89	1.56	0.86	0.89	1.56	0.84
Zimbabwe	0.97	1.69	0.88	0.96	1.55	0.87

n.a. Not applicable.

Note: The values in the table are the estimated coefficients from a regression using data pooled over all commodities (the sample covered 60 products, which vary by country) and over all years (1968–78).

Source: Authors' calculations, using data from FAO (various issues).

indicates that a high proportion of the variations in domestic prices are a
counted for by the variations in world prices.

There are no estimates of the exchange-rate elasticity for the United States
for Panama, Guatemala, and El Salvador, which used dollar-linked currencie
However, the values of the transmission elasticity for these countries are in li
with those obtained for the other countries.

The median of the exchange-rate elasticity, with these four countries ex
cluded, is 0.97, but the estimates vary greatly across countries. The variability i
the exchange-rate elasticity reflects the problem of determining the appropriat
measure of exchange rate for this analysis. In many countries this variable i
volatile because of inflation and changes in exchange-rate regimes. When th
exchange rate changes during the year, the rate that was applicable to a particu
lar commodity depends on the seasonality of that commodity and may diffe
from the variable used in the regression. A similar problem arises when ther
are multiple exchange rates, where whatever alternative is used represents a
compromise.

The effect of inflation on the results can be reduced by examining the identit
in equation 2 in terms of the real exchange rate:

$$(p_{it} - \bar{p}_t) = (p^*_{it} - \bar{p}^*_t) + (e_t + \bar{p}^*_t - \bar{p}_t) + s_{it} + u_{it} \quad (4)$$

where $\bar{p}_t$ and $\bar{p}^*_t$ are the logs of domestic and world price deflators, respectively
and the terms in parentheses represent real values. The estimation of equation 3
is repeated, with the real values replacing the nominal values. We deflate th
domestic and world prices by the domestic and U.S. consumer price index
respectively. The results appear in the last three columns of table 1. The result
for the regressions in real and nominal values should be the same under H_1 an
H_2. Indeed, the transmission elasticity is changed very slightly; its median valu
is 0.947. The median of the exchange-rate elasticity is 0.86, but for som
countries the estimate differs significantly from the respective nominal-valu
regressions, and the cross-country variability still exists.

Eliminating the effect of the exchange rate. Because we are interested largel
in the transmission elasticity, it is desirable to eliminate the effect of the ex
change rate. We consider two options. First, we compute within-year regression
(that is, regressions with year dummies). Such regressions use the price differ
ences between commodities for each year, and those do not reflect the exchang
rate. In this case, the regression equation takes the form

$$(p_{it} - p_{.t}) = \alpha + \beta(p^*_{it} - p^*_{.t}) + \epsilon_{it} \quad (5)$$

where $p_{.t} = \Sigma_i p_{it}/I$, the time-price average over commodities, with I being th
number of commodities. We use generic notations α, β, and ϵ for the intercept
the coefficient of world prices, and the disturbance term, respectively, althoug
their values are expected to vary from one equation to another. The prices i
equation 5 are also real, but unlike those in equation 4 they are deflated by thei

own sample averages. In this case there is no difference between the nominal and real variables. If we let $\bar{p}_t$ be the consumer price index used to convert nominal to real values, then the real domestic price to be used in equation 5 is $[(p_{it} - \bar{p}_t) - (p_{.t} - \bar{p}_t)] = (p_{it} - p_{.t})$, which is the nominal value. The same holds for real world prices. The results appear in the first column of table 2. The median value is 0.967, and on the whole the results are similar to those of the pooled regression with exchange rates included.

The second alternative is to express domestic prices in dollars:

$$(p_{it} - e_t) = \alpha + \beta p^*_{it} + \epsilon_{it}. \tag{6}$$

This approach was taken in an earlier version of this article, Mundlak and Larson (1990). The results of the estimation of equation 6 appear in the second column of table 2. They convey the same information as the previous regressions: the median value of the estimated transmission elasticity is 0.945.

Estimation results with all countries pooled together. The world price is the export unit value and, as such, it is not an average of domestic prices. After all, world trade constitutes only a small fraction of world production. To determine the extent to which the world price used here represents the domestic country price, the regression is estimated with all countries pooled together. In such an analysis the individual countries serve as repeated observations because they all face the same world price. The results are: 0.933 for R^2, 0.941 for the transmission elasticity, and 1.02 for the exchange-rate elasticity. The comparable estimates for real prices are 0.964, 0.943, and 0.980, respectively. It is thus concluded that the world prices used in the analysis are indeed representative of domestic prices.

The Policy Bias

The aforementioned regressions do not include a measure of the tax, s, as a variable because it is simply unobserved. This omission adds a component to the equation disturbance, and thus it reduces the R^2, which measures the importance of world prices in explaining the variations in domestic prices. More important, the omission may bias the transmission elasticity. It is often stated that countries pursue policies aimed at stabilizing domestic prices. Stabilization requires tax reductions when world prices are high and tax increases when world prices are low, which implies a negative correlation between world prices and taxes. Such a relationship is captured by

$$s_{it} = \pi_0 + \pi p^*_{it} + \nu_{it} \tag{7}$$

where ν is the error of this equation and $E(p^* \nu) = 0$. Combining equations 7 and 3, the regression equation for domestic prices is

$$p_{it} = (\alpha + \pi_0) + (\beta + \pi) p^*_{it} + \gamma e_t + \zeta_{it} \tag{8}$$

where $\zeta = \epsilon + \nu$. For convenience, we refer to π as the policy elasticity. Under

Table 2. *Estimated Transmission Elasticities from Regressions Excluding the Exchange Rate, 1968–78*

Country	*With year dummies*	*With domestic prices in dollars*	*With commodity means*[a]
Argentina	0.990	0.966	1.000
Australia	0.933	0.930	0.944
Austria	0.984	0.979	1.007
Bangladesh	0.710	0.715	0.731
Belgium and Luxembourg	0.973	0.972	0.993
Brazil	0.853	0.902	0.871
Burundi	0.884	0.862	0.901
Cameroon	0.881	0.890	0.900
Canada	1.018	0.999	1.033
Chile	0.970	0.878	0.785
Colombia	0.944	0.922	0.972
Costa Rica	0.931	0.908	0.944
Cyprus	0.934	0.925	0.948
Denmark	1.025	1.037	1.049
Ecuador	1.012	0.987	1.036
Egypt	1.231	1.208	1.271
El Salvador	0.904	0.903	0.925
Finland	1.000	0.967	1.026
France	0.953	0.949	0.968
Germany, Fed. Rep. of	0.995	0.989	1.037
Greece	0.906	0.912	0.925
Guatemala	0.971	0.907	0.972
India	0.774	0.737	0.794
Ireland	1.030	1.022	1.045
Israel	1.012	0.972	0.989
Italy	0.940	0.909	0.957
Japan	0.888	0.942	0.914
Kenya	1.090	1.064	1.112
Korea, Rep. of	0.904	0.926	0.907
Malawi	0.923	0.888	0.950

equation 2, $\beta = 1$ and the policy contributes to a discrepancy from 1. If we attribute all of the discrepancy of the estimated elasticity from 1 to the policy then a value of 0.95 for the transmission elasticity implies a value of -0.05 as an estimate for π, which is indeed very small.

Equation 7 assumes a uniform policy for all commodities and all years. This assumption is too strong and should therefore be weakened by generalizing the equation. This can be done by allowing commodity-specific policy elasticity denoted by π_i, and a tax level, denoted by π_{0t}:

$$s_{it} = \pi_{0t} + \pi_i p^*_{it} + v_{it} \tag{7a}$$

where $E(p^* v) = \text{cov}(\pi_{0i}, p^*_{it}) = \text{cov}(\pi_i, p^*_{it}) = 0$ for all t. The effect of this extension can be evaluated through the computation of between-commodity and

Table 2. *(continued)*

Country	With year dummies	With domestic prices in dollars	With commodity-means[a]
Malaysia	0.842	0.858	0.863
Mauritius	1.033	1.041	1.048
Mexico	1.040	0.985	1.058
Netherlands	0.989	0.985	1.016
New Zealand	1.051	1.029	1.068
Norway	0.976	0.977	1.006
Pakistan	0.804	0.744	0.829
Panama	0.970	0.937	0.971
Peru	0.852	0.868	0.902
Philippines	0.826	0.804	0.842
Portugal	0.970	0.959	0.982
South Africa	1.005	0.972	1.028
Spain	0.932	0.928	0.948
Sri Lanka	0.827	0.814	0.833
Sweden	0.955	0.930	0.986
Switzerland	1.018	1.039	1.043
Syria	0.977	0.978	1.002
Tanzania	0.989	0.977	1.012
Thailand	0.892	0.897	0.915
Trinidad	1.011	1.015	1.036
Turkey	0.943	0.952	0.961
United Kingdom	0.967	0.951	0.971
United States	0.958	1.005	0.958
Uruguay	0.800	0.796	0.809
Venezuela	0.933	0.910	0.963
Yugoslavia	1.020	1.011	1.041
Zambia	0.898	0.893	0.921
Zimbabwe	0.969	0.956	0.994

a. This column gives the values for the coefficient of world prices when using the between-commodity regression equation, in which the commodity price is an average over time and the world price is the only explanatory variable.

Source: Authors' calculations, using data from FAO (various issues).

within-commodity regressions. The appendix summarizes the formal relationships between the various estimators.

Between-commodity regressions. Averaging the variables over time we obtain

$$p_{i.} = (\alpha + \pi_{0i} + \gamma e_{.}) + (\beta + \pi_{i}) p^{*}_{i.} + \zeta_{i.} \tag{9}$$

where $p_{i.} = \Sigma_t p_{it} / T$, the commodity-price average over time, T is the number of years, and $e_{.} = \Sigma_t e_t / T$. The exchange rate is thus subsumed into the intercept and the between-commodity regression has only the world price as an explanatory variable. Its regression coefficient is

$$b(i) = \Sigma p_{i.}(p^{*}_{i.} - p^{*}_{..}) \;/\; \Sigma(p^{*}_{i.} - p^{*}_{..})^{2}. \tag{10}$$

Taking expectation using equation 9:

$$Eb(i) = \beta + \Delta(i) \tag{11}$$

and

$$\Delta(i) = \Sigma \lambda_i \pi_i \tag{12}$$

where $\lambda_i = (p^*_{i.} - p^*_{..})^2 / \Sigma(p^*_{i.} - p^*_{..})^2$ is the weight assigned to π_i. Thus $\Delta(i)$ is a weighted average of the commodity-specific policy elasticity, π_i. The value obtained for the coefficient of world prices, $b(i)$, appear in the third column of table 2. The median value is 0.971, so that it differs very little from that of the pooled regression. This implies that the weighted average of the policy elasticities is quite small.

Table 3. *Estimated Transmission and Exchange Rate Elasticities from within-Commodity Regressions, 1968–78*

Country	World prices (U.S. dollars)	Nominal values: World prices	Nominal values: Exchange rates	Real values: World prices	Real values: Exchange rates
Argentina	0.759	0.868	0.989	0.614	0.995
Australia	0.847	0.977	0.623	0.964	1.092
Austria	0.792	0.931	1.143	0.733	1.196
Bangladesh	0.630	1.094	0.604	0.911	0.938
Belgium and Luxembourg	0.828	0.921	1.145	0.750	1.154
Brazil	1.094	0.825	1.280	0.936	0.960
Burundi	0.579	0.585	1.466	0.257	1.233
Cameroon	0.873	0.903	1.057	0.671	1.062
Canada	0.797	0.962	−0.152	0.894	0.595
Chile	0.600	0.938	0.981	−0.010	1.376
Colombia	0.648	0.665	1.555	0.319	1.242
Costa Rica	0.655	0.500	2.367	0.404	1.308
Cyprus	0.826	0.820	−0.131	0.479	−0.142
Denmark	0.948	1.007	1.010	0.944	1.074
Ecuador	0.719	0.833	1.225	0.548	1.136
Egypt	0.964	0.600	−1.314	0.317	−1.901
El Salvador	0.759	0.759	2.413	0.618	1.408
Finland	0.636	0.769	2.263	0.713	1.582
France	0.846	0.939	1.305	0.817	1.255
Germany, Fed. Rep. of	0.748	0.979	1.252	0.935	1.369
Greece	0.845	0.746	1.451	0.667	1.164
Guatemala	0.697	0.972	0.000	0.869	1.350
India	0.437	0.272	2.783	0.218	1.487
Ireland	0.806	0.971	0.892	0.951	0.947
Israel	0.767	1.025	0.831	0.643	1.130
Italy	0.688	0.620	1.349	0.448	1.158
Japan	1.144	1.205	0.905	1.107	1.072
Kenya	0.750	0.634	1.739	0.532	1.107
Korea, Rep. of	1.006	0.846	1.190	0.767	1.106
Malawi	0.488	0.829	−1.162	n.a.	n.a.

Within-commodity regressions. The basic underlying equation for the within-commodity regression is obtained by subtracting equation 9 from equation 8, with π_{0i} and π_i replacing π_0 and π, respectively:

$$(13) \qquad p_{it} - p_{i.} = (\beta + \pi_i)(p^*_{it} - p^*_{i.}) + \gamma(e_t - e_.) + \zeta_{it} - \zeta_{i.}.$$

The within-commodity estimates were derived for the nominal and real versions with the exchange rate included and for the domestic variables in dollar prices. The results appear in table 3. The median of the transmission elasticities for the nominal prices, 0.937, is significantly higher than the corresponding values of 0.78 and 0.713 for the dollar prices and real prices, respectively.

To examine the source of this difference in the estimates, we simplify the

Table 3. *(continued)*

		Nominal values		*Real values*	
Country	*World prices (U.S. dollars)*	*World prices*	*Exchange rates*	*World prices*	*Exchange rates*
Malaysia	0.837	1.010	1.291	1.008	1.430
Mauritius	0.989	0.822	1.503	0.868	1.052
Mexico	0.646	0.904	1.091	0.452	1.131
Netherlands	0.819	0.992	1 208	0.973	1.245
New Zealand	0.764	0.939	0.207	0.883	0.958
Norway	0.801	0.974	1.381	0.880	1.388
Pakistan	0.367	1.285	0.144	0.531	1.287
Panama	0.604	0.969	0.000	0.633	−1.527
Peru	0.782	0.931	1.032	0.516	1.071
Philippines	0.597	1.078	0.578	0.465	1.222
Portugal	0.800	0.820	1.351	0.663	1.165
South Africa	0.626	0.922	0.325	0.721	1.972
Spain	0.816	0.822	1.227	0.682	1.108
Sri Lanka	0.686	1.205	0.367	1.004	0.895
Sweden	0.579	0.775	1.898	0.493	1.588
Switzerland	1.054	1.099	0.844	1.209	0.945
Syria	0.872	0.843	1.633	0.764	1.219
Tanzania	0.765	0.635	1.745	0.516	1.094
Thailand	0.769	0.781	1.206	0.506	1.087
Trinidad	0.866	0.887	1.784	0.771	1.272
Turkey	0.904	1.046	0.966	0.810	1.094
United Kingdom	0.781	0.943	0.671	0.864	0.518
United States	0.817	0.955	0.000	0.860	0.000
Uruguay	0.730	0.893	0.953	0.594	1.033
Venezuela	0.599	0.669	2.186	0.202	1.758
Yugoslavia	0.851	0.549	1.855	0.657	1.089
Zambia	0.713	1.003	0.988	0.736	1.940
Zimbabwe	0.697	0.969	1.280	0.832	1.191

n.a. Not applicable.
Source: Authors' calculations, using data from FAO (various issues).

exposition by ignoring the term with the exchange rate and writing the regression coefficient for commodity i as if it were a simple regression:[1]

$$b_i = \Sigma_t p_{it}(p^*_{it} - p^*_{i.}) \;/\; \Sigma_t(p^*_{it} - p^*_{i.})^2 \tag{14}$$

and express the within-commodity estimator, $w(i)$, as

$$\begin{aligned} w(i) &= \Sigma\Sigma p_{it}(p^*_{it} - p^*_{i.}) \;/\; \Sigma\Sigma(p^*_{it} - p^*_{i.})^2 \\ &= \Sigma_i b_i \omega_i \end{aligned} \tag{15}$$

where $\omega_i = \Sigma_t(p^*_{it} - p^*_{i.})^2 \;/\; \Sigma\Sigma(p^*_{it} - p^*_{i.})^2$. Thus the within-commodity estimator is a weighted average of the regression coefficients for commodity i. Taking expectation using equation 13:

$$Ew(i) = \beta + \Sigma_i \, \omega_i \pi_i. \tag{16}$$

The difference between equations 16 and 11 is in the way the commodity-specific slopes are taken into account. In equation 11 only one observation per commodity is used, and the exchange rate is eliminated from the equation. In equation 16 there are as many observations as years. If the difference of the transmission elasticities from 1 is considered as a weighted average of the commodity-specific policy elasticity, then the results suggest some variability in policy elasticity among commodities. We therefore examine this possibility more closely by presenting results for individual commodities and discussing possible sources of variations between the various estimates.

A similar analysis can be conducted for an alternative specification that allows for systematic variations of policy over time. This is discussed in Mundlak and Larson (1990), where it is shown that at the median the time effect on the policy elasticity, derived for domestic dollar prices, is -0.04. It seems that the variability over commodities is more important. To deal with the variability over time, a longer time series is needed.

Specific Commodities

Mundlak and Larson (1990) present transmission elasticities for wheat, coffee, and cocoa derived from equation 6, where the domestic prices are measured in dollars. The results are reproduced here in table 4. Wheat is chosen because it is often stated that staple foods are more susceptible to intervention that insulates domestic markets from world prices. Coffee and cocoa are internationally traded under cartel arrangements, and as such they may show a larger gap in the variations of domestic and world prices.

The median value of the transmission elasticity for wheat is approximately 0.65, with only 9 out of the 58 countries having a coefficient smaller than 0.5.

1. In practice b_i is obtained from the following regression:

$$p_{it} = a_i + b_i p^*_{it} + c_i e_t + \text{error}.$$

Equations 14 and 15 still apply when the prices are netted of the linear effect of e_t. This adjustment does not affect the interpretation.

The median value for coffee is 0.68; for cocoa it is above 0.84. The reason for concentrating on the estimates of equation 6 is that, with only 11 observations and a correlation between exchange rates and world prices, the results of the full equation 3 with an additional coefficient are less stable. This problem is overcome by pooling all countries together. The estimates of equation 3 for wheat with country-pooled data are 0.69, 1.03, and 0.969 for the transmission elasticity, exchange rate elasticity, and R^2, respectively. Interestingly, the value of the transmission elasticity is very close to the median value of table 4.

The policy elasticities for wheat, coffee, and cocoa are negative and larger in absolute value than those obtained for the pool of commodities. Still, for most countries, about 70 to 80 percent of the variations in world prices, depending on the commodity, are transmitted to domestic prices. Furthermore, the values of R^2 are on the whole quite high, indicating that world prices are the main source of variations in domestic prices.

The relationships between the within- and between-commodity estimates are illustrated in figure 1, where domestic dollar prices, measured in natural logs, are plotted against the log of world prices. Lines ac and bc represent regression lines with slopes $\beta + \pi_a$ and $\beta + \pi_b$ fitted to observations on two commodities, where both slopes are less than 1. Ellipsoids mark the clusters of observations

Figure 1. *Relationship between Within-Commodity and Between-Commodity Estimates*

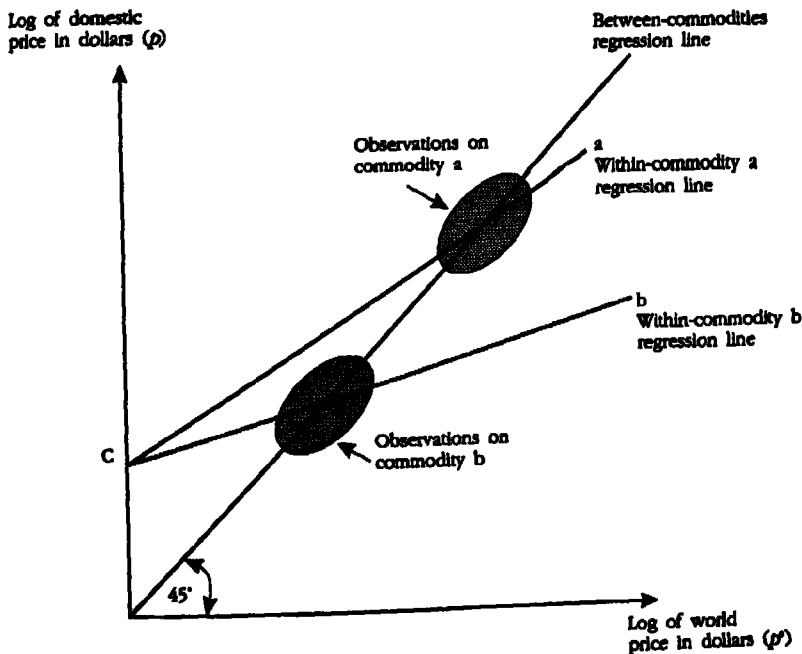



Table 4. *Estimated Transmission Elasticities for Wheat, Coffee, and Cocoa, 1968–78*

Country	Wheat		Coffee		Cocoa	
	Elasticity, *b*	R^2	Elasticity, *b*	R^2	Elasticity, *b*	R^2
Argentina	0.701	0.46	0.00	0.00	0.00	0.00
Australia	0.905	0.76	n.a.	n.a.	n.a.	n.a.
Austria	0.588	0.85	n.a.	n.a.	n.a.	n.a.
Bangladesh	0.655	0.37	n.a.	n.a.	n.a.	n.a.
Belgium and Luxembourg	0.626	0.81	n.a.	n.a.	n.a.	n.a.
Brazil	0.814	0.88	0.652	0.22	1.192	0.85
Burundi	0.518	0.31	0.680	0.99	n.a.	n.a.
Cameroon	0.152	0.68	0.530	0.88	0.615	0.86
Canada	0.954	0.66	n.a.	n.a.	n.a.	n.a.
Chile	0.836	0.29	n.a.	n.a.	n.a.	n.a.
Colombia	0.620	0.85	0.619	0.95	0.612	0.91
Costa Rica	0.554	0.93	0.940	0.99	1.075	0.94
Cyprus	0.477	0.86	n.a.	n.a.	n.a.	n.a.
Denmark	0.892	0.87	n.a.	n.a.	n.a.	n.a.
Ecuador	0.529	0.82	0.629	0.78	0.979	0.89
Egypt	0.562	0.86	n.a.	n.a.	n.a.	n.a.
El Salvador	0.621	0.78	1.054	0.93	0.927	0.69
Finland	0.413	0.83	n.a.	n.a.	n.a.	n.a.
France	0.582	0.85	n.a.	n.a.	n.a.	n.a.
Germany, Fed. Rep. of	0.646	0.87	n.a.	n.a.	n.a.	n.a.
Greece	0.715	0.83	n.a.	n.a.	n.a.	n.a.
Guatemala	0.687	0.83	0.862	0.96	0.975	0.91
India	0.405	0.85	0.142	0.39	n.a.	n.a.
Ireland	0.707	0.79	n.a.	n.a.	n.a.	n.a.
Israel	0.821	0.96	n.a.	n.a.	n.a.	n.a.
Italy	0.655	0.90	n.a.	n.a.	n.a.	n.a.
Japan	1.113	0.71	n.a.	n.a.	n.a.	n.a.
Kenya	0.780	0.81	1.006	0.99	n.a.	n.a.
Korea, Rep. of	0.903	0.81	n.a.	n.a.	n.a.	n.a.
Malawi	0.500	0.66	0.430	0.75	n.a.	n.a.

for the sample years for each commodity. The intercept of the price line of an individual commodity indicates a systematic difference between domestic and world prices or simply the level of distortion for the particular commodity. Thus we can have a slope of 1 and an intercept larger or smaller than 0, indicating a protection or tax, respectively. The lines are drawn with a common intercept so as to keep the same distortion rate. This is not essential, and other configurations are admissible.

The commodity-specific regressions estimate the slopes of these lines, and the within-commodity regression provides estimates of a weighted average of these commodity slopes. However, the between-commodity regression line is a statistical fit to commodity averages, labeled $\bar{p}$, and, unlike the within-estimator, the between-commodity estimator does not use the information represented in indi-

Table 4. *(continued)*

Country	Wheat: Elasticity, b	Wheat: R^2	Coffee: Elasticity, b	Coffee: R^2	Cocoa: Elasticity, b	Cocoa: R^2
Malaysia	1.008	0.97	0.837	0.76	0.844	0.94
Mauritius	0.692	0.82	n.a.	n.a.	n.a.	n.a.
Mexico	0.586	0.88	0.858	0.80	0.835	0.88
Netherlands	0.584	0.84	n.a.	n.a.	n.a.	n.a.
New Zealand	0.701	0.78	n.a.	n.a.	n.a.	n.a.
Norway	0.601	0.82	n.a.	n.a.	n.a.	n.a.
Pakistan	0.097	0.08	n.a.	n.a.	n.a.	n.a.
Panama	0.497	0.77	0.425	0.79	1.023	0.98
Peru	0.704	0.86	0.732	0.83	1.046	0.85
Philippines	0.609	0.85	1.018	0.87	0.754	0.92
Portugal	0.422	0.93	n.a.	n.a.	n.a.	n.a.
South Africa	0.454	0.93	n.a.	n.a.	n.a.	n.a.
Spain	0.546	0.90	n.a.	n.a.	n.a.	n.a.
Sri Lanka	0.586	0.64	0.809	0.71	1.085	0.82
Sweden	0.482	0.84	n.a.	n.a.	n.a.	n.a.
Switzerland	0.910	0.85	n.a.	n.a.	n.a.	n.a.
Syria	0.687	0.84	n.a.	n.a.	n.a.	n.a.
Tanzania	0.634	0.69	0.616	0.50	0.498	0.79
Thailand	0.995	0.84	0.461	0.86	n.a.	n.a.
Trinidad	0.729	0.81	0.604	0.75	0.702	0.97
Turkey	0.705	0.77	n.a.	n.a.	n.a.	n.a.
United Kingdom	0.708	0.77	n.a.	n.a.	n.a.	n.a.
United States	0.958	0.73	0.831	0.86	n.a.	n.a.
Uruguay	1.153	0.88	n.a.	n.a.	n.a.	n.a.
Venezuela	0.805	0.84	0.051	0.09	0.504	0.62
Yugoslavia	0.626	0.84	n.a.	n.a.	n.a.	n.a.
Zambia	1.187	0.96	0.715	0.52	n.a.	n.a.
Zimbabwe	0.624	0.84	0.434	0.75	n.a.	n.a.

n.a. Indicates not applicable.
Source: Authors' calculations, using data from FAO (various issues).

vidual commodity observations.

Because there is a large between-commodity spread in the prices, the slope of the between-commodity regression differs from the slopes of the individual commodities. An elasticity of 1 for the between-commodity estimate indicates that what is relatively expensive in the world market is also relatively expensive at home, or, more specifically, that the relative prices at home and abroad are, on average, the same. A slope smaller than 1 indicates that the more expensive the commodity is, the lower the tax rate or the larger the subsidy is.

Decomposition of the Pooled Regression Results

The pooled regression is a weighted average of the within- and between-commodity regressions (see the appendix), where the weights depend on the

variance components of the world prices. Table 5 presents a decomposition of the variations in world prices to the commodity and time components. The between-commodity variations dominate the within-commodity variations, and therefore estimates based on data pooled across time and commodities (first three columns of table 1) largely reflect the between-commodity variations. Thus the transmission elasticity of the pooled regression is close to the between-commodity value and close to 1 even though the commodity elasticities are smaller than 1.

A slope smaller than 1 for an individual commodity in this framework is consistent with a stabilization policy implemented by changing tax rates. But is a deviation from 1 for an individual commodity an exclusive outcome of policy? The answer is probably no. There are two important effects that are likely to

Table 5. *Sum of Squares of World Prices, 1968–78*

		Within			*Between*	
Country	*(1) (Total)*	*(2) Commodities (i)*	*(3) Time (t)*	*(4) Commodities and time (it)*	*(5) Commodities (i)*	*(6) Time (t)*
Argentina	562.8	81.3	493.0	11.5	481.5	69.8
Australia	511.0	75.1	446.6	10.6	436.0	64.5
Austria	426.4	57.2	373.2	7.7	370.4	53.4
Bangladesh	401.2	62.9	349.3	11.0	338.3	51.9
Belgium and Luxembourg	414.7	58.9	358.6	9.2	357.6	56.8
Brazil	576.0	91.8	499.3	17.5	483.6	77.0
Burundi	385.4	46.1	346.8	7.5	339.3	38.6
Cameroon	475.8	63.6	418.7	13.0	413.0	57.9
Canada	370.0	53.1	323.6	6.6	316.9	46.4
Chile	386.1	68.8	325.1	9.8	318.7	61.1
Colombia	538.3	82.3	469.5	13.5	456.1	68.8
Costa Rica	442.4	59.2	390.4	11.5	384.1	52.4
Cyprus	363.2	56.0	310.7	7.8	309.8	53.0
Denmark	297.1	48.5	249.5	7.9	250.9	48.3
Ecuador	542.9	83.4	473.6	14.1	459.5	69.3
Egypt	369.1	75.3	304.4	10.6	293.8	64.7
El Salvador	457.3	64.7	401.3	12.5	393.2	56.5
Finland	273.2	39.6	235.3	6.2	235.0	38.4
France	459.2	69.2	399.2	9.9	388.7	60.1
Germany, Fed. Rep. of	429.8	55.6	375.0	10.3	371.9	56.7
Greece	507.6	78.5	439.7	10.9	427.6	68.0
Guatemala	437.6	58.9	386.1	11.2	379.2	51.9
India	519.8	84.0	449.0	15.1	436.6	70.9
Ireland	296.2	35.1	268.1	5.8	268.5	28.1
Israel	404.6	72.5	338.3	10.9	330.5	67.1
Italy	453.3	84.5	377.8	11.0	370.0	75.6
Japan	562.8	78.4	494.6	12.6	486.5	68.3
Kenya	430.2	70.3	471.1	12.9	459.8	59.2
Korea, Rep. of	482.4	63.0	425.1	7.8	424.7	57.7
Malawi	375.8	50.1	334.7	9.0	325.8	41.1

bias the estimates downward: tradability and "measurement error."

Tradability. Although agricultural products are largely tradable, their domestic prices also reflect domestic inputs such as marketing, finance, storage, and transportation. (For an analysis of this subject, see Mundlak, Cavallo, and Domenech 1990.) To incorporate this extension, equation 2 is rewritten:

$$p_{it} = \tau_i(p^*_{it} + e_t + s_{it} + u_{it}) + (1 - \tau_i)p^d_t \tag{17}$$

where p^d_t is the natural log of the aggregate price of the domestic input, assumed to be the same for all products, and τ_i is the share of the tradable component in the price of commodity i. Under equation 17, the slope of the individual commodity price line is smaller than 1. The empirical transmission elasticity is now an estimate of $\tau_i + \pi_i$. As p^d_t is omitted from the regression, this estimate is

Table 5. *(continued)*

		Within			*Between*	
Country	(1) (Total)	(2) Commodities (i)	(3) Time (t)	(4) Commodities and time (it)	(5) Commodities (i)	(6) Time (t)
Malaysia	446.4	58.2	394.3	13.1	390.3	53.0
Mauritius	301.9	36.0	272.1	6.3	265.9	29.8
Mexico	522.9	92.6	440.5	14.8	433.2	82.8
Netherlands	322.3	56.1	269.1	7.5	268.7	53.6
New Zealand	400.1	51.5	355.3	6.7	348.6	44.8
Norway	274.5	38.8	236.1	6.9	234.6	39.6
Pakistan	411.6	75.2	346.2	11.6	337.2	65.5
Panama	320.5	40.2	286.7	8.4	280.5	34.0
Peru	556.7	90.0	483.0	16.9	462.6	74.3
Philippines	445.0	65.4	388.9	13.2	377.8	57 1
Portugal	516.6	69.9	453.9	12.2	448.0	63.2
South Africa	525.9	72.8	465.1	12.1	453.1	60.7
Spain	568.2	85.8	493.2	13.6	483.0	75.4
Sri Lanka	443.6	57.3	396.0	9.7	386.3	46.7
Sweden	365.1	53.2	310.5	10.7	315.1	56.4
Switzerland	351.6	46.5	310.8	6.4	301.3	41.0
Syria	397.4	71.8	335.9	9.7	325.6	61.7
Tanzania	535.0	78.2	469.3	14.5	457.3	65.8
Thailand	404.4	58.5	357.6	11.9	348.3	47.3
Trinidad	388.6	47.6	345.3	10.9	341.8	44.1
Turkey	455.6	78.6	387.2	10.2	377.0	68.4
United Kingdom	341.8	45.9	298.6	8.2	292.1	44.2
Uruguay	383.0	65.0	326.5	8.5	318.0	56.5
United States	528.0	81.5	457.8	13.2	445.8	70.5
Venezuela	420.8	64.3	367.0	11.3	355.0	53.9
Yugoslavia	488.1	76.7	420.6	9.2	411.4	67.5
Zambia	400.9	47.8	363.6	7.6	359.7	37.3
Zimbabwe	426.0	54.1	378.8	8.8	372.0	47.3

Source: Authors' calculations, using data from FAO (various issues).

biased. It is likely that the omitted variable, which is closely related to domestic inflation, is positively correlated with the exchange rate and that hence the bias is positive. Therefore, the deviation from 1 cannot be fully attributed to policy.

We can carry this analysis a step further and rewrite equation 17 as

$$p_{it} - p_t^d = \tau_i(p_{it}^* + e_t + s_{it} + u_{it}) - \tau_i p_t^d. \quad (18)$$

If we assume that the price of the domestic input can be approximated by the overall price level, then the dependent variable is real domestic price in the sense of equation 4. The difference is that in equation 18 the world price is nominal, whereas in equation 4 it is real, as well. If we ignore this difference, however, the estimates of equation 4 can be viewed as an approximation of the estimate of equation 18, with the last term omitted. But now the bias is negative, because the coefficient of p_t^d has a negative sign. Indeed, the estimates of the real regressions give somewhat smaller estimates for the transmission elasticities. The values of the within-commodity estimates are 0.937 for the nominal, 0.78 for the dollar prices, and 0.713 for the real. It is interesting to note that the effect of converting to dollars is similar to that of deflating by the overall price level. In terms of our discussion, the nominal estimate is biased upward, and the real is biased downward. Hence the difference from 1 obtained from the regression with real values can be viewed as an upper bound for an estimate of the sum of the share of the nontradable component and the policy elasticity, whereas the nominal regression provides a lower bound.

Measurement error. There are two good reasons to think about measurement error of a conceptual rather than mechanical nature. First, the basic equation 2 is applicable at a time when a trade takes place. Trade is not carried out continuously, however. Between trades, the world price is changing without necessarily affecting domestic prices. Stored commodities maintain an intertemporal arbitrage condition. As such, the spot prices respond to new information with respect to expected future world supply. Because arbitrage, either through trade or through storage, is costly, domestic prices, which are not backed by transactions, do not respond instantaneously to changes in world prices. Consequently, as illustrated effectively by Williams and Wright (1991), the dynamic paths of world prices and domestic prices within the year are likely to differ. When intrayear variations in the world price are summarized in equation 2 by a single figure, a discrepancy is built in between the domestic price and the pertinent world price. This is, of course, a short-term phenomenon, but it recurs with every new shock to the system. Because the prices are dated, this dynamic may matter and thus affect the results.

Second, the problem of deciding on the right deflators to convert the world prices from nominal to real is similar to the question of what exchange rate to use, which was discussed above. The deflators and exchange rates may bias the estimates downward. The bias may be substantial and may lead to a rejection of

the empirical validity of the law of one price. Also, in some countries (for example, Canada) the estimates of the exchange rate elasticities change very little over the sample period. Therefore the spread is not sufficient to get reliable estimates.

IV. The Data Set for the European Community

A potential problem of any empirical application is that the results emerging from the study reflect the idiosyncrasies of the way in which the data are collected, estimated, or reported rather than the underlying economic effects. This possibility carries a special weight in view of the doubt researchers express with respect to FAO data. Therefore we repeat the analysis on a separate data set for the European Community (EC), which has had an active agricultural policy as well as good data.

The EC agricultural policy is well financed and sophisticated in its execution and reporting mechanisms. Because it is well financed, any wedge between domestic and international prices can be expected to be longer lived than in lower-income countries. The data are taken from Herlihy and others (1989), and cover 25 years of producer prices. The commodity coverage available from the EC data is more limited than in the FAO data set, but it contains the major staple products. Table 6 presents some summary results. For the sample pooled across time and commodities, equation 3, the transmission elasticity varies between 0.91 and 1.01, and the values of R^2 vary between 0.84 and 0.92. The estimated exchange-rate elasticity, not reported in the table, varies between 0.79 and 1.06. The corresponding values for the pool of *all* the EC countries are 0.97, 0.97, and 1.03, respectively. The results for the between-commodities regressions are similar.

The within-commodity estimates of the transmission elasticity (equation 5) are somewhat smaller, with a median value of 0.74, compared with a median value of 0.96 for the between-commodity regression. This pattern is similar to what we observed above for the first sample. It is also similar to the results reported in Mundlak and Larson (1990) for equation 6, with domestic prices measured in dollars. We use this similarity to report in table 7 the results in Mundlak and Larson (1990) for individual commodities, with dollar domestic prices.

The results for the EC confirm the earlier results. Although the commodity coverage is different, the pooled elasticities for countries common to both samples are remarkably close. The estimates for wheat vary between 0.54 and 0.91, with a median at 0.70 and a value of 0.77 for the pool of all the EC countries. Recall that the median for the country estimates in table 4 is 0.66, and the estimates for the pooled country data derived from equation 3 is 0.69. For some of the other commodities, the median elasticities are also somewhat lower than the aggregates. But for milk the estimated transmission elasticity is larger than 1

Table 6. *Estimated Transmission Elasticities, European Community, 1960–85*

			Between				*Within*			
	Pooled		*Commodities*		*Time*		*Commodities*		*Time*	
Country	*Elasticity, b*	R^2	*Elasticity, b(i)*	R^2	*Elasticity, b(t)*	R^2	*Elasticity, w(i)*	R^2	*Elasticity, w(t)*	R^2
Belgium and Luxembourg	0.98	0.84	1.00	1.00	0.76	0.94	0.74	0.99	0.99	1.00
Denmark	0.96	0.88	0.95	1.00	1.09	0.98	1.05	0.98	0.94	1.00
France	1.01	0.88	1.04	1.00	0.82	0.98	0.78	0.99	1.03	0.99
Germany, Fed. Rep. of	0.96	0.85	0.97	0.99	0.54	0.88	0.47	0.99	0.96	0.99
Greece	1.00	0.92	1.02	1.00	0.68	0.99	0.65	0.99	1.02	1.00
Ireland	0.91	0.84	0.91	0.99	0.97	0.98	0.92	0.98	0.90	0.99
Italy	1.00	0.89	1.02	1.00	0.81	0.98	0.74	0.98	1.01	1.00
Netherlands	0.94	0.84	0.94	0.99	0.76	0.92	0.69	0.99	0.94	0.99
United Kingdom	0.95	0.90	0.96	0.99	0.90	0.95	0.86	0.98	0.96	0.99
All countries	0.97	0.97	1.02	1.00	0.89	0.98	0.85	0.99	1.00	1.00

Note: Exchange rate effects were included in estimates.
Source: Authors' calculations, based on data from Herlihy and others (1989).

)r several countries, indicating a strong adjustment of domestic prices that was ositively correlated with world prices.

V. Conclusions

By way of generalization, the deviation from unitary elasticity is, on the /hole, surprisingly small. The deviation from unitary elasticity is in part the esult of policy measures and in part the result of domestic inputs that are not ecessarily synchronized with world agricultural prices. This does not imply ıat policies generated with respect to particular products are not important in ffecting the prices of these products. They certainly affect the price levels, and, /henever a country taxes agriculture, the domestic prices will differ from world rices. Consequently, there are cross-country variations of prices. Such policies o not, however, prevent domestic prices from moving with world prices. Furnermore, world prices are the major contributor to variations in domestic rices.

In this analysis it was assumed that the world price is independent of the isturbances in the price equation. On the face of it, this assumption might be ɔo strong for the United States, and perhaps some other countries, when dealıg with some specific commodities. If this assumption were violated, world rice would be endogenous and the estimates would be subject to least-squares ias. However, this is not reflected in the results in any meaningful way.

What, then, is the role for domestic supply and demand? They determine the raded quantities of the traded goods, and the prices of the traded goods affect to large extent the prices of the specific factors in agriculture and thereby the upply of the nontraded goods. This is basically the mechanism of factor-price qualization. For instance, depressed world prices affect land prices, agricultural vage rates (through their effect on labor supply), and the price of quasi-fixed nputs. This spreads to all commodities.

An important implication for thinking about the dynamics of world agriculure (Mundlak 1989) is that we can think of the world as a closed economy acing a downward-sloping demand function that serves as a constraint to proluction growth. The trend in world prices is determined by the relative growth n world supply and demand. In this century supply has outpaced demand, and ıs a result real world agricultural prices have declined. The essence of our ınalysis is that such a decline should have taken place in all countries, regardless ıf whether their supply actually increased in relation to demand.

This implies that technical change and other permanent shocks that originate n one country but that are big enough to affect world prices eventually affect rices in all countries. Even though domestic policies affect prices, they cannot revent the covariations of domestic prices with world prices in the long run, ecause price distortion is costly, and public resources, like private resources, ıre finite. Passive countries, which are shock takers, should implement the ıecessary structural adjustments called for by the shock—including the enhance-

Table 7. *Estimated Transmission Elasticities for Selected Commodities, European Community, 1960–85*

	Belgium and Luxembourg		Denmark		France		Germany, Fed. Rep. of		Greece	
Commodity	Elasticity, b	R^2	Elasticity, b	R^2	Elasticity, b	R^2	Elasticity, b	R^2	Elasticity, b	R^2
Barley	0.714	0.873	0.972	0.940	0.691	0.865	0.684	0.858	0.632	0.873
Butter	0.522	0.822	1.101	0.930	0.607	0.915	0.736	0.907	0.716	0.916
Cattle	0.958	0.963	1.193	0.972	0.889	0.960	0.987	0.959	0.829	0.957
Cheese	0.921	0.947	1.287	0.976	0.844	0.972	1.022	0.989	0.773	0.920
Eggs	0.636	0.662	0.919	0.921	1.198	0.917	0.787	0.921	0.739	0.807
Maize	n.a.	n.a.	n.a.	n.a.	0.683	0.877	n.a.	n.a.	n.a.	n.a.
Milk	1.188	0.895	1.731	0.904	1.230	0.895	1.395	0.916	0.669	0.907
Oats	0.778	0.901	0.998	0.944	0.716	0.900	0.753	0.909	0.747	0.861
Pigs	0.722	0.938	0.819	0.956	0.571	0.910	0.754	0.950	n.a.	n.a.
Poultry	0.955	0.953	0.950	0.928	0.604	0.716	0.880	0.927	0.411	0.730
Potatoes	1.079	0.689	1.105	0.769	0.906	0.650	0.901	0.851	0.755	0.814
Rye	0.844	0.911	0.912	0.958	0.697	0.914	0.822	0.914	n.a.	n.a.
Sugar beets	0.741	0.712	0.885	0.825	0.798	0.777	0.718	0.766	0.645	0.649
Wheat	0.661	0.811	0.907	0.910	0.619	0.844	0.723	0.841	0.537	0.727

n.a. Not applicable.
Source: Authors' calculations, based on data from Herlihy and others (1989).

ment of technical change, if this is the source of the shock—rather than delay the process through taxation. This is certainly a very general statement, and it has to be properly interpreted when it comes to a particular policy; however, it is mentioned here in order to place possible implications of the analysis within a broader framework.

Appendix. A Summary of the Formal Relations between the Various Estimators

The analysis in the text differs somewhat from more familiar forms of panel data analysis. It is therefore useful to evaluate the results within a uniform framework. Let W, $B(i)$, $W(i)$, and $W(it)$ be projection (symmetric and idempotent) matrixes that generate residuals. They can be defined in terms of their operation on an arbitrary vector, x, of order IT: $Wx = (x_{it} - x_{..})$, $B(i)x = (x_{i.} - x_{..})$, $B(t)x = (x_{.t} - x_{..})$, $W(i)x = (x_{it} - x_{i.})$, $W(t)x = (x_{it} - x_{.t})$, and $W(it)x = (x_{it} - x_{i.} - x_{.t} + x_{..})$.

The parentheses contain the typical elements of the vectors in question. The following identities can then be derived.

$$W = W(i) + B(i) \qquad \text{(A-1)}$$

$$= W(t) + B(t) \qquad \text{(A-2)}$$

$$= W(i) + W(t) - W(it) \qquad \text{(A-3)}$$

$$= B(i) + B(t) + W(it). \qquad \text{(A-4)}$$

Table 7. *(continued)*

Ireland Elasticity, b	Ireland R^2	Italy Elasticity, b	Italy R^2	Netherlands Elasticity, b	Netherlands R^2	United Kingdom Elasticity, b	United Kingdom R^2	All countries Elasticity, b	All countries R^2
0.878	0.887	0.718	0.865	0.762	0.916	0.869	0.870	0.769	0.774
n.a.	n.a.	0.654	0.843	0.880	0.804	1.072	0.883	0.786	0.733
1.105	0.949	0.873	0.966	0.911	0.928	1.023	0.943	0.974	0.849
n.a.	n.a.	0.837	0.343	1.040	0.976	1.116	0.900	0.980	0.671
0.785	0.883	0.548	0.793	0.762	0.874	0.634	0.612	0.779	0.633
n.a.	n.a.	0.854	0.866	n.a.	n.a.	n.a.	n.a.	0.769	0.851
1.551	0.974	1.393	0.903	1.312	0.910	1.173	0.972	1.294	0.818
0.845	0.868	0.773	0.878	0.719	0.924	0.819	0.898	0.794	0.801
0.767	0.898	0.642	0.908	0.733	0.933	0.735	0.892	0.718	0.860
0.748	0.669	0.534	0.840	0.937	0.946	0.913	0.727	0.770	0.680
1.018	0.718	0.950	0.733	1.016	0.757	0.923	0.719	0.961	0.641
n.a.	n.a.	0.624	0.807	0.773	0.915	1.052	0.886	0.848	0.790
n.a.	n.a.	0.774	0.703	0.790	0.744	0.795	0.562	0.818	0.785
0.698	0.811	0.637	0.827	0.674	0.833	0.898	0.885	0.768	0.688

If p and p^* are the vectors of the two prices, then the regression coefficients of p or p^* can be presented in terms of $a = p^*Ap/p^*Ap^*$. When $A = W, B(i), B(t)$, the resulting estimators are b (pooled), $b(i)$ (between commodity), and $b(t)$ (between time), respectively. Also, when $A = W(i)$, $W(t)$, and $W(it)$, the coefficients are referred to as within commodity ($w[i]$), within time ($w[t]$), and within time and commodity $w([it])$, respectively.

We can then decompose the pooled regression coefficient:

$$b = p^*Wp/p^*Wp^* \quad \text{(A-5)}$$
$$= \theta w(i) + (1 - \theta)b(i)$$

where $\theta = p^*W(i)p^*/p^*Wp^*$ is the ratio of the within-commodity sum of squares and the total sums of squares, and the complement is a similar ratio for the between-commodity sum of squares. Table 5 presents a decomposition of the sum of squares of p^* by sources. Because p^* is the world price, the sums of squares should be the same for all countries. However, the set of commodities analyzed varies somewhat among countries, and therefore the numbers in the table differ accordingly. It is clear that the variations among commodities are greater by far than the variations over time; therefore, the pooled regression is closer to the between-commodity regression. A similar comparison can be made for the other estimators. Also, the results are easily generalized to multiple regressions, where the weights will be matrix, rather than scalar, weights (Mundlak 1978).

REFERENCES

The word "processed" describes informally reproduced works that may not be commonly available through libraries.

Ahmed, Raisuddin. 1988. "Pricing Principles and Public Intervention in Domestic Markets." In John Mellor and Raisuddin Ahmed, eds., *Agricultural Price Policy for Developing Countries*. Baltimore, Md: The Johns Hopkins University Press.

Anderson, Kym, Yujiro Hayami, and Masayoshi Honma. 1986. "Growth of Agricultural Protection." In Kym Anderson and others, eds., *The Political Economy of Agricultural Protection: The Experience of East Asia*. Sydney, Australia: Allen & Unwin.

Australia, Bureau of Agricultural Economics. 1985. *Agricultural Policies in the European Community: Their Origin, Nature, and Effects on Production and Trade.* Policy Monograph 2. Canberra: Australian Government Publishing Service.

Bale, Malcolm D., and Ernst Lutz. 1981. "Price Distortions in Agriculture and Their Effects: An International Comparison." *American Journal of Agricultural Economics* 63 (1): 8–22.

FAO (Food and Agriculture Organization of the United Nations). Various issues. *FAO Production Yearbook*. Rome.

Herlihy, Michael, Stephen Magiera, Richard Henry, and Kenneth Baily. 1989. *Agricultural Statistics of the European Community, 1960–85*. Statistical Bulletin 770. Washington, D.C.: U.S. Department of Agriculture.

Johnson, D. Gale. 1973. *World Agriculture in Disarray.* New York: Macmillan.

McCalla, Alex F. 1969. "Protectionism in International Agricultural Trade, 1850–1968." *Agricultural History* 43 (3, July): 329–44.

Mellor, John W., and Raisuddin Ahmed, eds. 1988. *Agricultural Price Policy for Developing Countries*. Baltimore, Md.: The John Hopkins University Press.

Mundlak, Yair. 1978. "On the Pooling of Time-Series and Cross-Section Data." *Econometrica* 46 (1, January): 69–86.

———. 1989. "Agricultural Growth and World Developments." In Alan Maunder and Alberto Valdés, eds., *Agriculture and Governments in an Interdependent World.* Proceedings of the Twentieth International Conference of Agricultural Economists, Dartmouth, Aldershot, England.

Mundlak, Yair, Domingo Cavallo, and Roberto Domenech. 1990. "Effects of Macroeconomic Policies on Sectoral Prices." *The World Bank Economic Review* 4 (1): 55–79.

Mundlak, Yair, and Donald F. Larson. 1990. "On the Relevance of World Agricultural Prices." WPS 383. World Bank, International Economics Department, Washington, D.C. Processed.

Valdés, Alberto, and Ammar Siamwalla. 1988. "Foreign Trade Regime, Exchange Rate Policy, and the Structure of Incentives." In John Mellor and Raisuddin Ahmed, eds., *Agricultural Price Policy for Developing Countries*. Baltimore, Md.: The Johns Hopkins University Press.

Williams, Jeffrey C., and Brian D. Wright. 1991. *Storage and Commodity Markets.* Cambridge, U.K.: Cambridge University Press.

World Bank. 1986. *World Development Report 1986*. New York: Oxford University Press.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO 3: 423-438

How Small Enterprises in Ghana Have Responded to Adjustment

William F. Steel and Leila M. Webster

Monitoring of adjustment has focused on larger, often state-owned enterprises, many of which have been adversely affected by the more competitive environment. Little is known, however, about the impact of adjustment policies on small firms. Firm-level data from Ghana show that the adjustment process was well under way across sectors and within firms and that new investment was taking place. The adjustment process had strained most firms' operations—profits were squeezed between rising input costs and weak domestic demand and low-cost competing imports. Small-scale industries were forced to become more competitive to survive. Interviews with owners of small firms revealed considerable entrepreneurial initiative in changing product mix and seeking newly opened market niches. Sample entrepreneurs fell into two broad groups: dynamic, successful adapters with good prospects (found mostly among small-scale enterprises) and stagnant producers who had not adapted to the new competitive environment (found mostly among microenterprises). For potentially dynamic small firms, the most critical constraint was lack of access to finance for working capital and new investment. Many microentrepreneurs were seriously constrained by a lack of purchasing power among the lower-income population and by saturation of the sector.

This article uses a survey of 82 manufacturing firms in Ghana to analyze how small, private enterprises responded to structural adjustment reforms and how their contribution to economic growth could be enhanced. Growth of private indigenous enterprises is important for sustained industrial development, given decreasing public ownership and uncertain prospects for foreign investment. But the absence of data on small firms, where most Ghanaian entrepreneurs are found, makes it difficult to assess their prospects in the changed environment. The survey reported here represents a rapid assessment technique for analyzing how different types of firms respond to adjustment policies and what constraints need to be addressed for them to realize their potential.

William F. Steel and Leila M. Webster are with the Industry and Energy Department of the World Bank. The data were gathered with the assistance of the Ghana National Board for Small-Scale Industries. The authors are grateful to E. K. Abaka, Frederick Gyebi Acquaye, and John Wayem for assistance and to Kwamena Adjaye, Surendra Agarwal, Benson Ateng, Nancy Barry, Stephanie Gerard, Don Mead, and two anonymous referees for their comments.

Ghana introduced an Economic Recovery Program (ERP) in 1983 to redress some of the causes of its long economic decline. The ERP's key elements had different implications for different types of industry. Supported by adjustment lending, import liberalization increased access to previously restricted inputs and spare parts (especially for small firms), but also broadened competition from imported products. Massive realignment of the highly overvalued exchange rate created new opportunities for export and import substitution, but adversely affected import-dependent industries by sharply raising the price of imported inputs and the cost of financing them. In principle, price liberalization permitted firms to pass through higher costs, but in practice many could not because demand was restrained by stabilization measures and competition was high, especially in self-employment activities with low barriers to entry.

The study focuses on small-scale enterprises (SSEs) for two principal reasons. The first reason is to investigate the view that African countries lack the dynamic entrepreneurs needed to initiate and sustain the process of industrialization. In this sense, SSEs—defined for this study as having from 4 to 29 full-time workers—can be distinguished from microenterprises with fewer than 4 workers, which are more likely to be informal and oriented toward income for survival. In Ghana SSEs had been discouraged by excessive regulation, distorted incentives, and state domination of industry. Hence we hypothesized that liberalization of markets would enable potential industrialists to emerge.

The second reason for our focus on SSEs is to assess the extent to which flexible organization and technology and relatively high labor intensity enable SSEs to survive during adjustment. We hypothesized that SSEs would be better able than large firms to respond quickly to the dramatic changes under Ghana's ERP and that smaller firms might absorb the surplus labor being shed by larger ones.

A consensus emerging from the literature is that the fundamental issue is *how* to create a policy and business environment that enables SSEs to contribute productively to industrial development, not *whether* SSEs have a role to play (Liedholm 1990; Schmitz 1982). Hence the study examines constraints on firms' ability to respond to changing incentives and how these differ by firm size. In particular, we hypothesized that microenterprises—defined in the Ghanaian context as having fewer than 4 full-time workers (including self-employment)—would serve more as a safety valve for surplus labor and be less able to respond to the ERP, given their low barriers to entry and limited access to resources.

The data were collected through interviews lasting one to two hours during three weeks in November 1989. The survey was conducted in two major urban centers, one medium-size town, and two rural towns to represent a cross-section of urban/rural and growing/stable settings. An effort was made to represent the range of activities in which SSEs are typically found, although the lack of suitable census data made it difficult to select a statistically representative sample. Sixty-six (80 percent) of the 82 firms surveyed fell into six subsectors that accounted for 81 percent of employment recorded in the industrial census (Ghana 1989)

and 86 percent of total manufacturing employment in the population census (Ghana 1987): food processing, textiles and garments, wood products, soap and cosmetics, metal products, and building materials. The oldest firm started in 1950 and the most recent in 1989. Forty-eight firms (59 percent) started before 1984, and 33 (40 percent) began operations between 1984 and 1989. These data were supplemented by an additional survey of selected questions for large firms with 30 or more workers.

The analysis focuses on differences across firm size categories, which generally yielded the most interesting results. Selected firm characteristics such as subsector and age are also analyzed. Firm characteristics are to some extent interrelated: no firms established since 1986 were large, and microenterprises were slightly overrepresented in the food and metal subsectors, as were SSEs in wood products. The small number of observations, however, precluded statistically disentangling the separate effects of different variables. The aim, rather, was to arrive at indicative answers to the research hypotheses.

Preceding the analysis of survey findings, section I summarizes the literature on the role of SSEs in industrial adjustment, and section II reviews the preadjustment context. Section III characterizes the entrepreneurs in the sample, and section IV examines changes in employment, output, and investment under the adjustment program. Sections V and VI analyze adaptive response and constraints on adjustment, and section VII presents the conclusions.

I. The Role of Small Enterprises in African Industrial Adjustment

In the period after independence, many African countries attempted to leap directly to a modern industrial structure through public investment in large-scale industries. The state often took the lead for lack of a strong indigenous entrepreneurial class and to avoid dependence on foreign investors. But inadequate attention to economic viability and market prospects resulted in substantial excess capacity, with many large firms unable to survive without heavy protection or subsidies. Many enterprises were squeezed, first, by economic crisis and, subsequently, by adjustment policies that reduced protection, cut back subsidies, restrained demand, and changed relative prices. Given budgetary restraints and a policy shift away from direct ownership of productive enterprises, governments have had to look increasingly to the private sector to take the lead in future industrialization.

Both cross-sectional and time-series data suggest that the industrialization process normally involves initial rapid growth of production in small-scale enterprises, some of which may expand into medium- and large-scale firms or survive in market niches even as large-scale industries gradually come to dominate the size distribution (Anderson 1982; Cortes and others 1987; Liedholm and Mead 1987; Little and others 1987; Nanjundan 1987; Staley and Morse 1965). In most African countries, the bulk of manufacturing employment is in self-employment and enterprises of fewer than 10 workers, while most output is

generally produced in large firms of more than 100 workers (Liedholm and Mead 1987). Thus, some observers argue that there is a "missing middle" in the size distribution of industry in African countries that must be filled for private initiative to sustain industrial development (Kilby 1988).

We follow much of the literature in distinguishing SSEs from household and cottage industries that are essentially self-employment with low barriers to entry. SSEs may be defined as production units with hired labor that involve enough investment in capital or skills to constitute a barrier to entry but are small enough to be managed by one person. Few of the firms surveyed were part of an informal sector in the sense of being outside the legal framework. Although many of the smaller firms appeared "informal" in the organization of their production process, all but a few were registered as businesses and paid taxes at least to the local district. Many gradations of "formality" in different aspects of business operation were observed. Hence we have avoided the term "informal sector" and instead have used size categories that can be objectively measured. The number of workers is a measurable and therefore convenient indicator of size, although it must be recognized that dividing lines between groups are arbitrary.

The following arguments have been used to justify the expansion of SSEs as a desirable strategy.

- Successful industrialization must have an indigenous base, and expansion of the SSE sector would help develop entrepreneurial and managerial skills as a basis for efficient indigenous investment in and management of larger industries (Bolton Committee 1971; Bruton 1990; Kilby 1988).
- Because SSEs tend to be relatively labor-intensive and to use low levels of technology, a strategy to expand the SSE sector is likely to be consistent with employment and income distribution objectives while allowing for sustained productivity increases through improvements in technology (Staley and Morse 1965; Steel and Takagi 1983).
- Abolishing restrictive licensing, tax incentives for large firms, and direct allocation of credit, materials, and foreign exchange to large firms (especially in the public sector) is likely to unleash investment by SSEs in previously untapped opportunities (Little and others 1987; UNIDO 1989).
- SSEs can respond flexibly under difficult and changing conditions because they do not depend heavily on infrastructure and because their typically low levels of technology allow product lines and inputs to be changed at relatively low cost (Morawetz 1974; Steel 1977).
- Even when large-scale industries dominate, many SSEs retain a competitive advantage by serving dispersed local markets, providing differentiated products with low-scale economies for niche markets, or specializing as subcontractors for larger firms (Anderson 1982; Anheier and Seibel 1987).

Adjustment policies such as those pursued by Ghana can be hypothesized to open opportunities for SSEs to expand. By introducing greater competitive forces

and opening up export opportunities, adjustment policies are intended to yield a more efficient, dynamic structure of production by forcing inefficient firms to lower production costs or die out. A substantial negative impact could be expected on inefficient large-scale firms that were created under heavy protection to import-intensive, import-substitution industries, which characterized much of Ghana's (and other African countries') industrial structure. However, this effect would be mitigated to the extent that firms could take advantage of increased availability of foreign exchange to utilize excess capacity.

SSEs could offset the negative effects of import liberalization to the extent that they could use local inputs to produce import substitutes or exports, thereby taking advantage of the incentives provided by a greatly devalued exchange rate. However, the ability of SSEs to grow would also depend on the ability of the financial system to shift resources to them from industries that are declining and on how severely demand and credit were restrained under stabilization policies.

II. Trends before Adjustment

As Ghana's economy declined in the late 1970s and early 1980s, foreign exchange scarcities curtailed production in its large import-substitution industries, which depended heavily on imported inputs. Capacity utilization fell from 40 percent in 1978 to 21 percent in 1982. Although public policy prevented employment from falling in state-owned enterprises, large private firms recorded in the industrial statistics in Ghana reduced employment by a third during this period.

Through innovative use of local raw and waste materials, some technically astute small entrepreneurs were able to fill gaps left by declining imports and large-scale output by producing import substitutes such as soap, metal products, and vehicle spare parts (Anheier and Seibel 1987; Dawson 1990). However, many SSEs were constrained by lack of access to complementary imported inputs, such as chemicals for tie-dye, screws for carpenters, ink for printers, perfumes for cosmetics producers, and even yarn for weavers of traditional kente cloth. There is no conclusive evidence as to whether opportunities for SSEs outweighed constraints during the period of economic decline and tight controls.

Stagnant income per capita in the 1960s stimulated growth in self-employment as people attempted to maintain family incomes. Women accounted for 63 percent of the total increase in nonagricultural employment from 1960 to 1970—half of them in manufacturing and 91 percent of them in enterprises of fewer than 10 workers (Steel 1981). This process accelerated with steeply declining real incomes in the late 1970s and early 1980s. Female participation in the formal labor force jumped from 64 percent in 1970 to 82 percent in 1984, and women accounted for the entire increase in manufacturing employment during that period (Steel 1981; Ghana 1987). With rising inflation in the 1970s and 1980s, many modern sector workers turned to self-employment ac-

tivities to supplement their declining real wages, while others were forced to work on their own when they were laid off. As a residual source of employment and economic survival, the microenterprise sector is generally thought to have expanded in the face of a declining economy.

III. ENTREPRENEURS DURING ADJUSTMENT

The survey identified small enterprises as an important locus of investment under the adjustment program. Almost half of the firms with fewer than 30 workers were established after the 1983 reforms. In contrast, none of those with 30 or more workers was established after 1983 (including 31 respondents to a related survey of large-scale firms). Although it is not known what percentage of new firms would be "normal" for each size group, the difference in the proportion of new businesses between the small and large groups is striking.

Many new owners were highly entrepreneurial in that they saw a profitable opportunity and took it. None followed their parents into business, as against 38 percent of pre-1984 owners. Their primary motivation was to apply their training (43 percent versus 13 percent of pre-1984 owners). These new entrepreneurs were better educated than their predecessors, both formally and through significant past job experience. The number of new small firm owners who finished middle school was twice that of pre-1984 owners (60 percent versus 29 percent).

The most successful SSE entrepreneurs moved into profitable niches producing specialized, nontraditional items, such as freezers, water coolers, and drums, to compete with increasingly expensive imports. Some undertook innovative processing of locally available materials, such as fuel briquettes from sawdust and knives from used band-saw blades and metal packing strips. Some left large companies to start their own business. Most started small and moved up, some more quickly than others.

In contrast, new microentrepreneurs (with fewer than 4 workers) were more likely to have entered business through the apprenticeship route: 44 percent versus 25 percent of pre-1984 microentrepreneurs. Ghana's informal apprenticeship system ensures that existing skills are maintained in trades such as garments, carpentry, vehicle repair, and metal-working; however, this also ensures that the numbers of young adults with similar skills will increase, without necessarily imparting the entrepreneurial and innovative abilities needed to become more productive. For example, one seamstress compensated for declining demand by increasing her paying apprentices to 37. With negligible orders for garments, the girls learned to sew on paper bags. The seamstress solved her immediate need for income but trained her future competition. The apprentices' futures probably will depend more on the growth of income among Ghana's workers than on their abilities as seamstresses.

Women owned 15 of the 82 firms surveyed; all but 2 of them had fewer than 10 workers and were found in food processing, textiles, and garments. These

subsectors accounted for 78 percent of total female manufacturing employment in 1984 (Ghana 1987). Although female entrepreneurs generally were less educated than their male counterparts in the sample, among the microenterprises women tended to be better educated than men (43 percent with 12 or more years of schooling as against 28 percent of men) and to have been in business much longer (an average of 17 years). One interpretation is that many women may view successful self-employment as a satisfactory, permanent source of income rather than as a step toward expansion. Another is that women have limited opportunities to expand their enterprises because of other substantial demands on their time. The problems women face in business in Ghana appear to arise mostly from their concentration in easy-entry, highly competitive activities and from gender-related differences in education and family responsibilities.

IV. Employment, Output, and Investment under Adjustment

This section analyzes how adjustment affected employment, production, and investment at the firm level and pays particular attention to differences between firms established before and after the ERP was launched. The evidence indicated that adjustment policies shifted the advantage away from larger firms and that entry and growth were particularly high among SSEs.

Changes in Employment

Table 1 shows the average annual growth in employment by firm size during 1975–83 and 1983–89 for sample firms established by 1975. During 1975–83, incomes fell and import controls were tightened as foreign exchange became increasingly scarce. Nevertheless, employment grew at 7.6 percent a year in the medium- and large-scale firms in the sample, presumably because imports were channeled to these firms and because the government maintained strong pressure on both public and private sector firms not to lay off workers. Microen-

Table 1. *Employment Growth, by Age and Size of Firm, 1975–83 and 1983–89*
(weighted average annual percentage growth)

	Firms established by 1975		*Growth since start*		*Number of firms established by*		
Firm size	*1975–83*	*1983–89*	*Established by 1983*	*Established during 1984–89*	*1975*	*1983*	*1989*
1–3 workers	0.0	16.0	9.7	7.6	4	17	33
4–9 workers	−0.2	6.2	1.3	45.1	7	16	26
10–29 workers	0.5	3.1	8.1	19.1	9	9	16
30+ workers	7.6	−17.2	1.1	n.a.	5	7	7

n.a. Not applicable.

Note: Size categories and weights are based on total employment in 1983 for firms established by then, to reveal the impact according to firm size at the *beginning* of adjustment. For firms established from 1984 to 1988, size categories are based on 1989 employment and weights on employment at start-up.

Source: Survey data.

terprises and SSEs, however, lacked government protection and suffered stagnant or declining employment as incomes and their access to inputs fell.

The preceding data are for individual firms in the sample. In the aggregate, as some firms shut down, production and employment were falling in the large-scale (especially private) sector. Conversely, aggregate employment (although not necessarily production) in microenterprises and SSEs is thought to have increased, as people sought alternative and supplementary sources of income.

The picture reversed sharply under the ERP from 1983 to 1989. Employment fell rapidly in the medium- and large-scale sample firms, partly because many were able to shed excess labor and partly because they were squeezed between high costs of imported inputs and greater competition from liberalized imports. Employment rose in all the other size categories, in part reflecting expansion as smaller firms gained greater access to inputs and rural demand recovered.

One striking feature of the recovery was that a much higher proportion of the firms surveyed reported an increase in employment (62 percent) than in production (39 percent; table 2). This was true for all size groups and most subsectors; employment growth lagged only in food products and soap and cosmetics. Increased labor absorption in part represents substitution in response to drastically eroded real wages and increased interest rates, making labor relatively cheap. A sharp increase in the number of microenterprises confirms that many workers turned to this sector for additional income (part-time workers accounted for 50 percent of microenterprise employment, as against an average of 13 percent for all firms surveyed).

Table 2. *Production and Employment Trends, by Size of Firm, for Firms Established by 1983 and during 1984–89*
(percentage of respondents in each category)

Impact	*All firms surveyed*[a]	*Number of workers in firms established by 1983*				*Number of workers in firms established during 1984–89*				*Large firm survey*
		1–3	*4–9*	*10–29*	*30+*	*1–3*	*4–9*	*10–29*	*30+*	
Change in production										
Increase	39	29	31	33	43	34	70	43	n.a.	58
Decrease	43	65	44	67	43	46	20	14	n.a.	29
Change in employment										
Increase	62	47	56	78	71	56	80	71	n.a.	35
Decrease	16	18	25	11	29	6	0	29	n.a.	39
Number of firms	82	17	16	9	7	16	10	7	0	31

n.a. Not applicable.

Note: Size categories are based on employment at the time of the survey in 1989. Compared with table 1, the microenterprise category includes some unsuccessful firms that declined, and the largest category includes some successful firms that grew. Firms reporting no change are not shown; so figures may not add to 100 percent.

a. Excluding large-firm survey.

Source: Survey data.

Production and Competition

Between 1983 and 1989, approximately the same proportion of firms in the total sample increased production (39 percent) as decreased (43 percent; table 2). A clear pattern emerges when firms are broken down by size and period of establishment. Firms established before the ERP had a relatively high propensity to decline under adjustment policies. In contrast, firms in all size groups established since 1983 were more likely to have increased their output than those already in existence. New firms with 4 to 29 workers (SSEs) had especially high ratios of growing to contracting firms. This growth in the newer firms may be partly attributable to the initial spurt that characterizes many SSEs in their first three years (Liedholm 1990), as well as to their choice of activities with high growth potential. Nevertheless, microenterprises with 1 to 3 workers were more likely to have decreased than increased output.

Among firms with fewer than 30 employees, three-quarters reported that other SSEs were the primary source of competition. Although 61 percent claimed that competition had increased since 1983, only 12 percent mentioned imports as a major source of that competition. The principal subsectors affected by import competition were metal products (21 percent of firms, mostly in agricultural machinery, which could be imported duty-free) and soap and cosmetics (29 percent of the subsector, especially cosmetics firms that did little more than repackage imported materials).

Investment

Contrary to some observers' perceptions, investment was taking place, at least among small firms. Nearly half of the microenterprises sampled were established since 1983 (high birth and death rates are expected in this group); while 38 and 44 percent of firms with 4 to 9 and 10 to 29 workers, respectively, entered since 1983. As import costs rose, many of these new firms sprang up to supply low-cost substitutes, particularly from local materials. Examples included pottery, locally mixed paints, and simple agricultural implements. In addition, about half of the firms established by 1983 had purchased some new equipment by 1989 (table 3).

V. ADAPTATIONS

Many firms adapted to new price incentives by altering their product lines—an important motivation for replacing outmoded equipment that had deteriorated during the late 1970s and early 1980s. The product mix was altered during the ERP by 34 percent of all firms surveyed and generally by higher proportions for those established by 1983 (table 3). As the construction industry picked up, for example, metal workers began producing metal gates and burglar alarm systems. Unable to compete with imports, a chalk producer shifted to producing starch for the textile industry. Changes in product mix were especially

Table 3. *Production and Input Changes under Adjustment, by Size of Firm, for Firms Established by 1983, 1983–89*
(percentage of respondents in each category)

Change	All firms surveyed[a]	Number of workers in firms established by 1983			
		1–3	4–9	10–29	30+
Bought new equipment	49	43	56	56	57
Product mix changed	34[b]	40	27	63	45[b]
Selling in different markets	15	12	19	0	43
Have exported (directly or indirectly)	9	6	8	7	14
Have considered exporting	38	28	46	44	43
Imported share of raw materials					
Actual percentage share	34.0[b]	19.1	19.1	62.8	52.[illegible]
Change since 1983					
Greater	9[b]	11	11	25	8[b]
Smaller	28[b]	22	11	13	83[b]
Easier to get					
Imported inputs	73[b]	89	64	50	88[b]
Domestic inputs	57[b]	64	44	17	67[b]
Harder to get or too costly					
Imported inputs	12[b]	0	18	33	0[b]
Domestic inputs	20[b]	21	25	33	0[b]
Number of firms	82	17	16	9	7

a. Includes firms established after 1983 as well as before.
b. Includes an additional 24 firms from a separate survey of large-scale enterprises.
Source: Survey data.

prevalent in the soap and cosmetics and the textiles and garments subsectors (83 and 50 percent of firms, respectively). Those manufacturing food, wood, and metal products had less need to change because their dependence on imported inputs was low (19, 10, and 18 percent, respectively) and their products were adapted to local tastes, making them less vulnerable to competition from standardized imports.

Along with changes in product mix, a number of firms shifted their marketing strategies. Large firms in particular had to seek out new domestic markets (43 percent; table 3). Devaluation raised the price of exports, and the response was especially high in the timber industry. New, small sawmills were buying logs from those with timber concessions and sawing them into lumber for larger exporting firms. Of the total sample, 38 percent were interested in exporting, even though only 9 percent had any direct or indirect experience with exports. SSE exports tended to be somewhat haphazard and on a small scale. For example, a tie-dye producer sold a relatively small volume of her fabric to a German buyer, and a producer of commercial freezers sold several units to traders from Nigeria and Mali.

Use of Inputs

The use of imported inputs was affected in two opposite ways by Ghana's adjustment policies. First, import liberalization made imports more available to

small firms, previously excluded by the large-scale firms that were favored by the import licensing system. Second, devaluation raised the cost of imported inputs, favoring industries with high proportions of domestic inputs and value added. Among firms in the sample, the same proportion used imported inputs exclusively as used only domestic raw materials (22 percent). On the average, the proportion of imported raw materials used by firms with fewer than 10 workers (19 percent) was about a third that used by larger ones, making them less vulnerable to devaluation (table 3). While the majority of respondents of all sizes reported that imports had become easier to obtain, only for those with 10 to 29 workers did increased access outweigh the substitution effect: twice as many (25 percent) increased their import content as decreased (13 percent). None of the firms in wood, metal products, and textiles and garments reported any change in import content (partly because import shares were already relatively low).

The overwhelming majority of firms with 30 or more workers reported substitution of domestic for imported inputs under the ERP (83 percent, in contrast to an average of only 16 percent of firms with fewer than 30 workers; table 3), although no significant change in average import content appears in recorded statistics for large-scale industry from 1984 to 1988. One reason for the relatively low substitution found among smaller firms was that relative prices may not have changed much in favor of domestic materials, most of which are tradable. A much larger share of SSEs cited the high price of local raw materials as a problem than that of imported inputs (for which they previously had to pay black market prices). The price of agricultural raw materials had risen because food prices were liberalized. The price of cloth to the tie-dye and garment subsector remained high because extra tariff protection was provided to large textile mills, and costs in the greycloth factory were not significantly reduced. Since timber is exportable, its domestic price tended to rise pari passu with devaluation. Good scrap metal became increasingly scarce. Thus there may have been supply-side constraints to rapid expansion of some SSEs that were based on local raw materials.

VI. Constraints on Adjustment and Growth

Although some firms found market niches with good growth potential, many were not able to pass on fully the increased costs of raw materials and equipment to consumers in the form of higher prices because of weak demand and an inflow of competing imports. Most firms experienced a financial squeeze, including those on sound financial footing in the past. The survey revealed that if demand were eliminated as a problem, finance would be an even more severe constraint on growth. In sum, the incentive side of the adjustment process was working—less efficient firms were being squeezed—but the financial side was not functioning adequately to enable more efficient firms to grow.

Respondents to the survey identified lack of access to credit (especially for

working capital), demand, and increased cost of inputs as their principal operating problems, but the relative importance varied by size. The importance of credit increased with firm size, while demand problems were most evident among microenterprises. SSEs were more concerned about the high price of local inputs, whereas larger firms placed more emphasis on imported input costs. SSEs also were most eager to replace old equipment and hence were most affected by lack of access to term lending.

The majority of SSEs (58 to 64 percent of firms with 4 to 29 workers) saw their ability to respond to market incentives as constrained primarily by lack of access to institutional credit for working capital and for equipment (table 4). The other two main constraints—the high cost of local materials and the need to replace old equipment—indicated why credit was needed to facilitate their supply re-

Table 4. *Constraints by Size of Firm, 1983–89*
(percentage of respondents in each category)

Constraint	*All firms surveyed*	*Number of workers* 1–3	4–9	10–29	30+
Main constraint on sales					
Resources (could sell more)	52	44	58	64	43
Demand					
Can sell current production but no more	24	19	23	21	57
Can't sell current production	24	38	19	14	0
Credit harder to get	84	83	100	100	100
Bank loan					
Have ever had one	32	18	19	60	86
Have tried to get one in past 5 years	47	28	48	73	71
Competition is greater	61	53	64	33	71
Major competition from					
Imports	21[a]	13	8	7	50[a]
Small firms	65[a]	75	72	72	37[a]
Top four operating problems[b]					
Credit					
For raw materials	57	55	54	63	71
To buy equipment	29	18	35	50	14
Demand					
Consumers lack of money	35	55	23	19	29
Too many other firms	17	21	12	25	0
Can't afford my product	11	15	12	6	0
Inputs					
Local material prices	22	30	23	69	14
Imported input prices	18	9	23	25	29
Can't get local materials	11	15	15	0	0
Equipment					
Needs replacing	23	12	35	31	14
Number of firms	82	33	26	16	7

a. Includes an additional 24 firms from a separate survey of large-scale enterprises.
b. Firms could list up to four problems, so the percentages can add to more than 100 percent. Problems listed by less than 11 percent of firms are not shown.
Source: Survey data.

sponse. Many of these firms had orders for their products but insufficient working capital to meet them, due in part to slim profit margins and delays in collecting from customers. Because responses were based on owners' perceptions, internal management problems were likely to be underrepresented as constraints. Nevertheless, almost all the owners of firms with fewer than 10 workers expressed a desire to improve their business skills, particularly those relating to technology and accounting.

Microenterprises were much more likely to be constrained by demand: 38 percent could not even sell their current production (table 4). They attributed weak demand primarily to people's lack of money—reflecting decreased real income per capita since the 1970s as well as restrained demand under the ERP. Increased competition from other small firms was also a factor: a growing number of competitors were slicing up a diminished pie. Thus the fortunes of the smallest firms depend most on growth of demand among low-income consumers (rural and urban), who are their principal markets.

In contrast to firms with fewer than 30 workers, larger firms saw imports as their main competition, and none cited competition from other domestic firms among their top four problems. Although they did not feel as immediately constrained by demand as microenterprises, 57 percent thought that demand could not absorb further increases in their production (table 4). These results confirm the importance of import liberalization to force large firms to act competitively.

Credit

As could be expected under tight monetary policies that curtail government spending and restrain credit, almost all of the survey respondents agreed that credit had become *harder* to get under the ERP (table 4). The survey also indicated significantly less access to bank loans for firms with fewer than 10 workers (fewer than 20 percent had ever had a loan) than for those with 10 to 29 and more than 30 workers (60 and 86 percent, respectively). More than 70 percent of firms with 10 or more workers had applied for a loan since the ERP began, thus indicating a high demand for credit. Only 9 percent of firms said they would refuse a loan for working capital at 30 percent interest, and only 25 percent (mainly microenterprises) would decline credit for investment purposes. Microenterprises often demurred because they doubted that they could sell enough to repay their loans.

Regulatory Environment

Given the severity of finance and demand problems, the regulatory environment was rarely mentioned among the top four problems, except for taxes (cited by 14 percent of all respondents and 30 percent of large-scale firms). Infrastructure (especially electricity interruptions) and transportation costs were cited by 5 to 7 percent. Only 5 percent mentioned the business environment generally, and just 3 percent (large-scale firms only) listed regulations or licensing as a problem.

Regulatory problems appeared to affect larger firms more than smaller ones, and exporters more than those producing for the domestic market. Location was the only significant regulatory concern of firms with fewer than 10 workers, many of whom had been affected by relocation and demolition efforts by urban authorities.

Nevertheless, the business environment may be more of a concern for new investors than for existing producers who have already adapted to it. Uncertainty about the economy was seen as a restraint on new investment by 38 percent of respondents, especially those with 10 or more workers. More than a third of large firms also expressed some reservations about the government's attitude toward private investment.

VII. Conclusions

Structural adjustment policies have begun altering the structure of industrial production in Ghana. Changes in the exchange rate, trade policy, and price controls have had different effects on incentives and profits. Although both positive and negative effects were found within each size and subsector group, differences between group averages indicate that structural changes were occurring. Large firms were particularly affected by competition from liberalized imports, and the increased cost of imported inputs forced many to increase the content of local raw material. Small enterprises were especially responsive in adapting their product lines to changing market opportunities and in taking advantage of increased access to imported inputs. Many entrepreneurs sought new products, techniques, and markets, and others would do so if they had greater access to resources, especially finance. New small enterprises were generally performing well.

Nevertheless, the scarcity of credit and the absence of mechanisms to shift resources from declining firms to those with greater growth potential appeared to restrain investment to take advantage of new opportunities. SSEs were quick to identify lack of credit for raw materials and equipment as major constraints to expansion, and they expressed strong interest in loans at current interest rates. Increased access to credit would greatly facilitate SSEs' contributions to recovery and adjustment of the industrial sector.

In contrast, microenterprises were much more likely to view weak demand as a major constraint—a consequence of the sharp fall in income per capita in the decade before the ERP and of restraints on demand under the ERP. Employment in microenterprises (including self-employment) mushroomed as formal sector employment and incomes diminished, but much of this growth was driven by excess supply of labor rather than by production-generating demand. Income and productivity may decline further without increased demand for the type of goods and services produced by this already overcrowded, low-wage sector. Policies that put more money into the hands of the low-income population (for example, through improving the terms of trade for farmers) are likely to have

greater effect on microentrepreneurs' earnings than supply-side measures to assist individuals.

As demand and credit problems are resolved, taxation and the business climate may come increasingly into play as constraints on investment, at least for larger firms; smaller firms tend to be less concerned with regulations. Economic and political stability and a positive attitude toward private profit would help reduce uncertainty—an important consideration for new investors. The cost of doing business could be reduced through lower, more transparent taxation. Export procedures, as well as financing, need improvement for SSEs to live up to their export potential. In addition to such measures, a long-run strategy should include complementary education and technical training to enhance the ability of Ghana's small entrepreneurs to contribute to industrial development.

References

The word "processed" describes informally reproduced works that may not be commonly available through libraries.

Anderson, Dennis. 1982. "Small Industry in Developing Countries: A Discussion of Issues." *World Development* 10 (11, November):913–48.

Anheier, Helmut K., and Hans Dieter Seibel. 1987. *Small-Scale Industries and Economic Development in Ghana: Business Behaviour and Strategies in Informal Sector Economies*. Cologne Development Studies 3. Saarbrucken, Germany: Breitenbach.

Bolton Committee. 1971. *Bolton Committee Report: Small Firms*. Cmnd. 4811. London: Her Majesty's Stationery Office. Quoted in Ian M. D. Little and others, 1987. *Small Manufacturing Enterprises*. New York: Oxford University Press.

Bruton, Henry. 1990. "Broad-Based Growth." Paper presented at a workshop on Research Priorities for Policy Reforms Supporting Broad-Based Growth and Democracy. USAID/APRE/SMIE. Washington, D.C. Processed.

Cortes, Mariluz, Albert Berry, and Ashfaq Ishaq. 1987. *Success in Small- and Medium-Scale Enterprises: The Evidence from Colombia*. New York: Oxford University Press.

Dawson, Jonathan. 1990. "The Wider Context: The Importance of the Macroenvironment for Small Enterprise Development." *Small Enterprise Development* 1 (3, September):39–46.

Ghana, Republic of. 1987. *1984 Population Census of Ghana: Demographic and Economic Characteristics, Total Country*. Accra: Ghana Statistical Service.

———. 1989. *Ghana National Industrial Census 1987*. Accra: Ghana Statistical Service.

Kilby, Peter. 1988. "Breaking the Entrepreneurial Bottleneck in Late-Developing Countries: Is There a Useful Role for Government?" *Journal of Development Planning* 18:221–49.

Liedholm, Carl. 1990. *The Dynamics of Small-Scale Industry in Africa and the Role of Policy*. GEMINI Working Paper 2. Washington, D.C.: USAID.

Liedholm, Carl, and Don Mead. 1987. "Small-Scale Industries in Developing Countries: Empirical Evidence and Policy Implications." International Development Paper 9. Michigan State University, Department of Agricultural Economics, East Lansing. Processed.

Little, Ian M. D., Dipak Mazumdar, and John M. Page, Jr. 1987. *Small Manufacturing Enterprises: A Comparative Analysis of India and Other Economies*. New York: Oxford University Press.

Morawetz, David. 1974. "Employment Implications of Industrialization in Developing Countries: A Survey." *Economic Journal* 84 (335, September):491–542.

Nanjundan, S. 1987. "Small and Medium Enterprises: Some Basic Development Issues." *Industry and Development* 20:1–50.

Schmitz, Hubert. 1982. "Growth Constraints on Small-Scale Manufacturing in Developing Countries: A Critical Review." *World Development* 10 (6, June):429–50.

Staley, Eugene, and Richard Morse. 1965. *Modern Small Industry for Developing Countries*. New York: McGraw-Hill.

Steel, William F. 1977. *Small-Scale Employment and Production in Developing Countries: Evidence from Ghana*. New York: Praeger.

———. 1981. "Female and Small-Scale Employment under Modernization in Ghana." *Economic Development and Cultural Change* 30 (1, October):153–67.

Steel, William F., and Yasuoki Takagi. 1983. "Small Enterprise Development and the Employment-Output Trade-Off." *Oxford Economic Papers* 35:423–46.

UNIDO. 1989. "Environment Conducive to Sustained Growth of Small- and Medium-Scale Enterprises." Discussion Paper ID/WG.492/4 for the First Consultation on Small- and Medium-Scale Enterprises including Co-operatives (Bari, Italy), Vienna. Processed.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 3: 439-458

The Dynamics of Optimal Gradual Stabilizations

Alex Cukierman and Nissan Liviatan

Inflation inertia may be quite tenacious because of the simultaneous interaction between policy actions and inflationary expectations under imperfect credibility. This result is particularly relevant for understanding some of the failed efforts to stabilize inflation in South America. This article deals with the issue of inertia in the framework of imperfect information about the type of the policymaker and extends the existing models to an infinite horizon. Because policymakers do not have perfect control of inflation, a "frivolous stabilizer" may deviate from the policies of a "serious stabilizer" without necessarily being unmasked immediately. When the difference in the ability of "strong" and "weak" policymakers to control inflation is large, unexpected inflation may be persistently negative for quite a while, thus causing reduced economic activity and giving the indication that credibility is low. If the policymaker persists with the stabilization, this pattern gradually disappears as his reputation rises. But before this final stage the serious policymaker has to compromise his inflation objective in view of adverse expectations about his type and pay the cost of imperfect credibility.

In some countries, particularly in Latin America, inflation has been quite tenacious in spite of recurring attempts at stabilization. Many of these attempts probably failed because the stabilization packages did not include a serious commitment to slow down the rate of growth of the money supply and to reduce the deficit. But even when such a commitment was in place, as in Chile and Argentina during the mid-1970s, inflation came down rather slowly and was accompanied by substantial and sustained reductions in the level of economic activity.

It is obviously possible to claim that if the monetary brakes had been applied more strongly, stabilization would have been faster. It is not clear, however, that such a course of action would have yielded better overall results. The policymaker may find it preferable to stabilize gradually because of credibility problems, thus providing a rigorous foundation for "inflation inertia" (Kiguel and Liviatan 1988).

Alex Cukierman is with the Department of Economics at Tel Aviv University, and Nissan Liviatan is with the Department of Economics at Hebrew University and the Country Economics Department at the World Bank. A previous version of the paper was presented at the September 1991 European meeting of the Econometric Society, in Cambridge, U.K. The authors would like to thank Ariel Rubinstein, John Driffill, and several anonymous referees.

The behavior of the policymaker while in office depends on the public's expectations about inflation (Barro and Gordon 1983). The public's expectations reflect, in turn, the possibility that the policymaker is not serious about achieving stabilization (not a "serious stabilizer"). When the public has little faith that the policymaker has really become more concerned about price stability, even a serious stabilizer does not necessarily reduce inflation quickly in order to avoid very substantial decreases in economic activity. Hence, when a serious stabilizer is in office, inflationary expectations as reflected in nominal contracts turn out to be too high ex post, which causes them to be revised and reduced. This reduction in inflationary expectations raises the policymaker's reputation as a serious stabilizer. But this learning process, although optimal, is gradual because actual inflation is affected not only by the deliberate decisions of the policymaker but also by unpredictable events over which the policymaker has no control. Because subsequent policy actions also depend on expectations, and because expectations are reduced slowly, even a serious stabilizer may be led to stabilize gradually.

If in office for a sufficiently long period of time, the serious stabilizer will ultimately build up a good reputation and deliver a much lower rate of inflation. But the period of stabilization will be protracted, and during most of it unexpected inflation will be negative, thus creating a persistent lull in economic activity. This lull occurs in spite of the fact that the serious stabilizer partially accommodates inflationary expectations in order to minimize the combined costs of low employment and high inflation.

Much has been made of inflation inertia arising from long-term, overlapping wage contracts and backward-looking, formal or informal indexation arrangements. This article demonstrates that even without these sources of inertia, inflation may be quite tenacious because of the simultaneous interaction between policy actions and inflationary expectations when there is imperfect credibility. Policy actions respond to expectations, which in turn are updated optimally, but sluggishly, in light of actual inflation.

Practically all previous models in which the public is uncertain about what type of policymaker is in office postulate a finite horizon for the policymaker (Backus and Driffill 1985a, 1985b; Barro 1986; Vickers 1986; Persson and van Wijnbergen 1987; Andersen and Risager 1987, 1988; Cukierman and Liviatan 1991a). In this article we characterize the path of inflation and other variables when the policymaker has an infinite horizon. This extension makes it possible to determine whether a serious stabilizer is eventually able to deliver the best performance.

Because policymakers do not have perfect control over inflation, a "frivolous stabilizer" may deviate from the policies that would have been followed by a serious stabilizer without necessarily being revealed as a frivolous stabilizer immediately. Imperfect control of inflation makes it more costly for a serious stabilizer to build up credibility quickly.

The first dynamic implication of our framework is that inflation comes *down*

during a gradual stabilization process, in contrast to models with mixed strategies of the Backus and Driffill (1985a) or Barro (1986) type, in which inflation *goes up* during this process. The second is that as inflation comes down, so do inflationary expectations. In Barro and in Andersen and Risager, inflationary expectations are fixed over time. Moreover, because of their fixed horizon, these models imply that a serious stabilizer may not reap the benefits of better performance until the last few periods, or even the very last period, of the game. Hence, the fixed-horizon assumption of these models does not give a serious stabilizer who cares somewhat about economic activity much incentive to pursue stabilization. An infinite horizon thus seems necessary in order to understand situations in which it is optimal to embark on a stabilizing path even when the formation of reputation is lengthy and gradual.

The analytical framework we develop provides a natural vehicle for the discussion of the effect of foreign aid on the credibility of stabilization. Untied aid, by permitting a higher level of economic activity, reduces the incentive to inflate and increases the credibility of stabilization. Conditional foreign aid reinforces these tendencies and speeds up the process of stabilization even further.

Section I presents a simple, infinite-horizon, Barro-Gordon (1983) type of framework, with uncertainty about the type of policymaker and with imperfect control of inflation. Section II introduces the public's (Bayesian) process of learning. A full characterization of the model's solution for the case in which a serious stabilizer opts for a gradual stabilization is developed in section III. Section IV presents the dynamic features of the equilibrium solution and uses it to interpret the stabilization in the United Kingdom under Margaret Thatcher and the stabilization in Chile. Section V discusses the effect of conditional and unconditional foreign aid on the credibility of stabilization.

I. The Model and Its Rationale

The model is designed to bring out the public's uncertainty concerning the likelihood that the policymaker is serious about stabilizing inflation and the effect of this uncertainty on policy. In order to capture the public's uncertainty, we assume that there are two types of policymakers: weak and strong. Both types of policymakers dislike inflation and desire to maintain employment above the natural level of employment. (The natural level of employment is the level at which expected and actual inflation rates are equal.) The desired level of employment of the "weak" policymaker is higher than that of the "strong" policymaker. More precisely, the combined cost of inflation and of being below the desired level of employment for each policymaker in period t is

$$z_t = \frac{A}{2}\,(N_i^* - N_t)^2 + \frac{\pi_t^2}{2}, \qquad i = w,s \tag{1}$$

where z denotes cost; A is the relative preference for price stability versus employment objectives and is a parameter common to both types of policymaker;

N_i^* is the desired level of employment of policymaker i; w denotes the weak policymaker and s the strong policymaker; N_t is the actual level of employment; and π_t is the actual rate of inflation.

Neither policymaker has perfect control over the rate of inflation. In particular, when a policymaker plans to generate inflation at rate π_{pt}^i, the actual rate of inflation is

$$\pi_t^i = \pi_{pt}^i + \epsilon_t^i, \qquad i = w,s \tag{2}$$

where π_{pt}^i is the actual inflation under policymaker i, and ϵ_t^i is a stochastic noise term whose variance is inversely related to the degree of control of the policymaker over the rate of inflation. Imprecise control of inflation is the result of either imperfect control of the money supply (Cukierman 1992, chapter 9) or the policymaker's uncertainty about money demand. The public cannot determine with certainty to what extent a change in monetary expansion occurs because of the policymaker's deliberate plans and to what extent it occurs because of an error in the policymaker's forecast of money demand (Cukierman 1992, chapter 13; Canzoneri 1985). The change obviously may result from a combination of both possibilities. We assume that the stochastic noise term has zero expected value and is distributed uniformly. In particular,

$$p_r\,[\epsilon_t^w = x] = \begin{Bmatrix} 1/2a_w & -a_w \le x \le a_w \\ 0 & \text{otherwise} \end{Bmatrix} \tag{3a}$$

$$p_r\,[\epsilon_t^s = x] = \begin{Bmatrix} 1/2a_s & -a_s \le x \le a_s \\ 0 & \text{otherwise} \end{Bmatrix} \tag{3b}$$

where x is a particular realization of the noise term, and a_w and a_s denote measures of the imprecision of inflation control by the weak policymaker and strong policymaker, respectively. Assuming that the weak policymaker is less precise in controlling inflation than the strong policymaker is, $a_w > a_s$.

The policymaker can affect employment by creating inflation that was unanticipated at the time nominal contracts were concluded. This situation leads to a conventional expectations-augmented Phillips relation that is summarized in equation 4:

$$N_t - N_n = a(\pi_t - \pi_t^e), \qquad a > 0 \tag{4}$$

where N_n denotes the natural level of employment, and π_t^e represents the rate of inflation expected at contracting time for the period of the contract. Equation 4 states that the deviation of employment from its natural level is positively related to unanticipated inflation. Substituting equation 4 into equation 1 and setting a = 1 for simplicity,

$$z_t \equiv z(d_i, \pi_t, \pi_t^e) = \frac{A}{2}[d_i - (\pi_t - \pi_t^e)]^2 + \frac{\pi_t^2}{2} \tag{5}$$

where

$$d_i \equiv N_i^* - N_n, \qquad i = w,s. \tag{6}$$

Thus d_i is the (positive) divergence between the level of employment desired by a policymaker of type i and the natural level of employment.

From the Barro-Gordon (1983) analysis, when the public is fully informed about the type of policymaker holding office and there is perfect control of the money supply (ϵ_t^i being identically equal to zero), the equilibrium rate of inflation is

$$\pi^i = Ad_i, \qquad i = w,s. \tag{7}$$

This is the time-consistent or subgame-perfect equilibrium. It is obtained by letting the policymaker choose the level of inflation (π) so as to minimize the costs in equation 5. The policymaker would do this by taking the rate of inflation expected at contracting time (π_t^e) as given and then imposing rational expectation. Under perfect information, rational expectation amounts to the requirement $\pi_t^e = \pi_t$ (Cukierman 1992, chapter 3).

Because the difference between the desired rate of employment and the natural rate of employment is positive, the equilibrium rate of inflation is positive. Moreover, because there is no uncertainty of any kind, wage setters fully anticipate the subsequent action of the policymaker. As a consequence, employment is always at the natural level in spite of the fact that inflation is positive. Obviously, the same level of employment could have been obtained with zero inflation, provided wage setters had believed the policymaker would choose zero inflation. But the wage setters have no reason to hold such a view, because zero inflation is not optimal for the policymaker after the wage contract has been made. This dynamic inconsistency of monetary policy induces a suboptimally high rate of inflation (Kydland and Prescott 1977).

Because the difference between the desired level of employment and the natural level of employment is greater for the weak policymaker than for the strong one, the equilibrium in equation 7 implies that the weak policymaker produces a rate of inflation that is higher than the rate of inflation produced by the strong policymaker. The intuition is that the weak policymaker is known to have a larger employment objective and is rightly expected to inflate at a higher rate. Hence, wage setters demand higher wage increases than in the case in which the strong policymaker is known to be in office.

When stabilization programs are introduced, the public is usually uncertain about their outcome. We model this uncertainty by assuming that either the weak or the strong policymaker is in office forever but that the public is not sure which type is in office. As time passes, the public learns from the realizations of inflation which type is likely to be in office. But this process may be protracted: because both types of policymakers have imperfect control of inflation, realizations of past inflation do not necessarily convey precise information about the type to the public. The assumption that one type of policymaker is in office forever is obviously not made for its realism. Policymakers do change, and the relative emphasis on employment versus price stability may change even within the same administration. The assumption is made to illustrate the potential difficulties that low credibility brings to a strong policymaker even in the favor-

able case in which a particular type of policymaker is in office forever and public uncertainty concerns only his or her type.

The policymaker's plans about how to handle inflation are made by taking both the current and the future values of the costs $z(\cdot)$ into consideration. In particular, when in office, either type of policymaker makes plans for current and future inflation rates that will, given the information available at the time, minimize the expected present value of costs. This present value, as of the present period (denoted by 0), is given by

$$(8) \qquad E_{p0}\sum_{t=0}^{\infty}\beta^{t} z(d_i,\pi_t,\pi_t^e), \qquad 0 < \beta \leq 1, i = w,s$$

where β is a discount factor that measures the policymaker's rate of time preference and E_{p0} is an expected value conditioned on the information available to the policymaker when the policymaker picks the average rate of inflation of the initial period. The weak policymaker is more sensitive to the costs of low employment in all periods because the difference between the desired and the natural level of employment is greater for the weak policymaker than for the strong one.

The length of a period is determined by the length of nominal wage contracts. Within each period, nominal wage contracts are made on the basis of the expected rate of inflation between the previous and the current period. Then the policymaker picks the planned rate of inflation for the period, taking those expectations (or nominal contracts) as given. This sequence shows that the government cannot precommit itself to a level of inflation. Actual inflation for the period is determined, through equation 2, by the policymaker's decision and by the realization of the uncontrollable inflationary shocks. The public observes the actual rate of inflation before it sets inflationary expectations and nominal contracts at the beginning of the next period. However, the public never observes the two components of the actual rate of inflation—one component planned by the policymaker and the other uncontrollable—separately.

II. Stabilization and the Evolution of Reputation

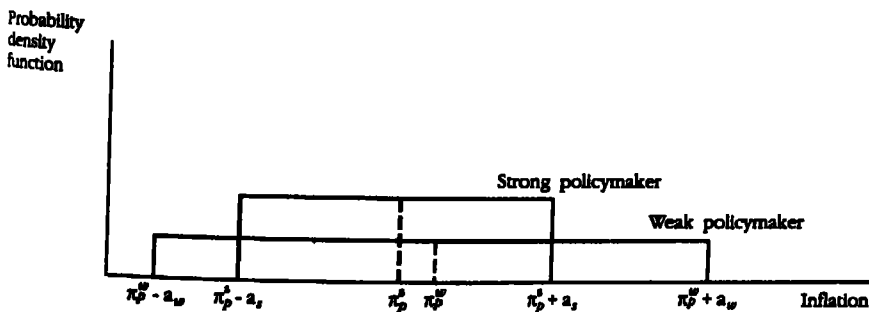
Inflation may be quite high for a while because a weak policymaker has been in office. If a strong policymaker then settles permanently in office and announces that a stabilization phase has commenced, the public will remain skeptical. This is because the earlier, weak policymaker would have had an incentive to make similar statements. We model the public's skepticism by assuming that its prior probability that the announced stabilization has been made by a strong policymaker is a number that is strictly bounded between zero and one. The smaller the prior probability, the lower the initial "reputation" of the policymaker (Backus and Driffill 1985a, 1985b; Vickers 1986).

Inflation uncertainty normally rises with the level of inflation (Engle 1982;

chapter 18 of Cukierman 1992). We capture this feature in a simple manner by assuming that the weak policymaker, who always plans to inflate at a higher rate than does the strong policymaker, also does not control inflation as tightly ($a_w > a_s$).

In spite of this difference in the precision of inflation control, the public may not be able to ascertain with certainty, even many periods after the policymaker has taken office, that a strong type is in office. The reason is that imprecise control of inflation by both policymakers prevents the public from clearly separating one type from the other, even if, as is normally the case, they plan to produce different average rates of inflation. Figure 1 illustrates why this is so for arbitrary equilibrium strategies of the two types of policymakers. From equations 2 and 3a, actual inflation when a weak policymaker is in office is between $\pi_p^w - a_w$ and $\pi_p^w + a_w$. Similarly, actual inflation when a strong policymaker is in office is between $\pi_p^s - a_s$ and $\pi_p^s + a_s$. When actual inflation is in the common range, which is the range of the strong policymaker in figure 1, there is no way for the public to clearly identify that a strong policymaker is in office. But because the probability that a rate of inflation in the common range has been produced by a strong type is larger than the probability that it has been produced by the weak type (as can be seen from figure 1), realizations of inflation in the common range raise the reputation of the policymaker in office.

This intuitive argument is confirmed by Bayes' formula, which shows how reputation evolves when the realization of inflation is in the common range. Given the actual rate of inflation in the current period, Bayes' formula relates the public's subjective probability that the policymaker in office is strong to the probability that the policymaker was strong before the realization of current inflation. Equation 9 shows the updating formula when actual inflation falls in the common range.

Figure 1. *Strategies of Different Policymaker Types and the Corresponding Distributions of Inflation*

Note: The relative positions of the two distributions in the figure are meant to be illustrative. Other relative configurations, some of which are discussed in the text and illustrated in figure 2, are possible.

$$(9) \quad \alpha_{t+1} \equiv Pr[i = s \mid \pi_t]$$

$$= \frac{Pr[\pi_t \mid i = s]\, Pr[i = s]}{Pr[\pi_t \mid i = s]\, Pr[i = s] + Pr[\pi_t \mid i = w]\, Pr[i = w]}$$

$$= \frac{\frac{1}{2a_s}\alpha_t}{\frac{1}{2a_s}\alpha_t + \frac{1}{2a_w}(1 - \alpha_t)} = \frac{\alpha_t}{\alpha_t + \frac{a_s}{a_w}(1 - \alpha_t)}$$

where Pr is an abbreviation for probability. Because the ratio of the imprecision of inflation control of the strong policymaker to that of the weak one is smaller than one, the probability that the policymaker is strong is greater in the current period (after the realization of π_t) than in the previous period for all t as long as inflation falls in the common range of figure 1. In other words, given that the public is still uncertain about the type in office, an additional realization of inflation in the common range always raises the reputation of the policymaker in office. (If the ratio of imprecision of inflation control is one, then realizations of inflation in the common range are not informative.)

When the weak policymaker is in office, there is a positive probability that inflation will fall in the noncommon range. When inflation actually falls in this range, the policymaker is revealed as being weak with certainty, and the public's prior probability that the policymaker is strong jumps to zero. The public knows that such a realization of inflation could not have occurred had a strong policymaker been in office. Thus, given the strategies postulated in figure 1 for the weak and strong policymaker types, the weak type will ultimately be revealed as weak, although that may take a long time. However, it will never be demonstrated to the public with full certainty that the strong type is in office when this is the case. But the reputation of the strong type will increase monotonically, reaching one asymptotically as time goes to infinity.

III. Equilibrium Strategies, Policy Convergence, and Expectations under Gradualism

The policymakers' equilibrium choice of strategy determines the probability that the public will be able to identify the policymaker as weak or strong. When the divergence between the equilibrium strategies of the two types of policymakers is small compared with the divergence in the precisions of their control over inflation, the probability that the public will be able to separate a strong policymaker from a weak policymaker may be zero. Conversely, the probability of the public's being able to make a distinction is usually positive, and may even be one, if the divergence between the equilibrium strategies is sufficiently large in relation to the difference in the precisions of inflation control.

The case of a small divergence in equilibrium strategies leads to equilibria in

which it is optimal for a strong policymaker to stabilize gradually. The case of a large divergence corresponds to situations in which it is optimal for a strong policymaker to take a chance with a "shock treatment." Which of the two methods of stabilization is optimal generally depends on the policymakers' reputations at stabilization, their rate of time preference, the precision of their control over inflation, and other parameters. The main focus of this article is on equilibrium when the optimal method of stabilization is gradual.

Characterization of Gradual Stabilizations

This section fully characterizes the equilibrium strategies of the two types of policymakers, but under a (provisional) assumption. The assumption is that the range of possible inflation rates produced by the weak policymaker's strategy fully covers the range of inflation rates that could be produced by the strong policymaker's strategy. In this case, the dynamic optimization problem decomposes into a series of one-period maximization problems. A convenient feature of the uniform distributions postulated for the noise terms (ϵ_t^i) is that all observations on inflation are equally informative and are independent of the magnitude of inflation as long as they all fall within the common range. Hence, as long as the strategies (planned inflation rates) of the weak and strong policymakers are such that the range of possible rates produced by the strong policymaker is fully covered by the range of rates that could be produced by the weak policymaker, the probability that the type will be revealed is independent of the precise location of the planned inflation rates within this range. As a consequence, within this range, either policymaker can select a current strategy to maximize the value of current objectives without paying attention to future values of the objectives.

Because equilibrium strategies depend on the process of forming expectations and this process depends in turn on what the public knows about these strategies, expectations and equilibrium strategies are determined simultaneously. The appendix shows that the equilibrium strategies are given by equation 10 or, alternatively, by equations 11 and 12, where B is a positive combination of parameters (whose precise form appears in the appendix). Expectations are given by equation 13, which is also derived in the appendix.

$$(10) \qquad \pi_{pt}^i = \frac{A}{1+A}(d_i + \pi_t^e), \qquad i = w,s$$

$$(11) \qquad \pi_{pt}^s = Ad_s + (1-\alpha_t)B(d_w - d_s)$$

$$(12) \qquad \pi_{pt}^w = Ad_w - \alpha_t B(d_w - d_s)$$

$$(13) \qquad \pi_t^e = \alpha_t(Ad_s) + (1-\alpha_t)(Ad_w).$$

Equation 13 implies that inflationary expectations in period t are a weighted average of the discretionary rates of inflation that would have been chosen by each policymaker under perfect information. Because the difference between the

desired and natural level of employment is greater for the weak policymaker than for the strong policymaker ($d_w - d_s > 0$), equations 11 and 12 imply that imperfect information causes both policymakers to have equilibrium strategies that converge toward each other compared with the equilibrium strategies they would have under perfect information. When the value of the reputation indicator is one-half ($\alpha_t = 1/2$), the planned rate of inflation of the weak policymaker decreases and that of the strong policymaker increases by the same amount compared with their full-information counterparts. As the value of the reputation indicator increases above one-half, the tendency of the weak policymaker to move toward the strong one increases, and the tendency of the latter to move toward its weak counterpart diminishes. The converse happens when the value of the reputation indicator decreases below one-half.

Because the marginal costs of being away from the desired level of employment are higher the further away actual employment is from its desired level (equation 1), both types of policymakers partially accommodate inflationary expectations. This is reflected by the positive, but smaller-than-one, coefficient $[A/(A + 1)]$ of the rate of inflation expected at contracting time (π_t^e) in equation 10. Because the expected rate of inflation is affected by what the public knows about the equilibrium strategies of both policymakers, the expected rate is somewhere between the rates of inflation that the strong and the weak policymakers plan to generate. Because the latter rate is larger, it follows that the weak policymaker's planned rate of inflation will be lower than it would have been under perfect information. The strong policymaker's planned rate of inflation is higher than it would have been under perfect information. The public's uncertainty pulls the policies of the two types toward each other. As reputation increases, expectations approach the equilibrium strategy of the strong policymaker. Hence, the tendency of the strong policymaker's planned rate of inflation to converge toward that of the weak policymaker diminishes, and the tendency of the weak policymaker's planned rate of inflation to converge toward that of the strong policymaker increases. Alternatively, as reputation diminishes, the tendency of the strong policymaker to compromise on the full-information strategy increases. As the reputation indicator approaches zero, the strong policymaker's planned rate of inflation approaches $Ad_s + B(d_w - d_s)$ (equation 11). The difference between this strategy and the strong policymaker's strategy under perfect information tends toward $B(d_w - d_s)$. Thus, with poor credibility, the actual policy of a strong policymaker may very well resemble that of a weak policymaker.

To assure that the strategies given in equations 10, 11, and 12 are indeed equilibrium strategies, it is necessary to assure that neither the strong nor the weak policymaker wants to deviate from the range in which the rates of inflation that could have been produced by the weak policymaker fully cover the range of rates that could have been produced by the strong policymaker. A sufficient condition for the weak policymaker not to want to deviate from this range is

$$a_s > \pi_{pt}^w - \pi_{pt}^s \text{ for all } t, \tag{14}$$

which, using equations 6 and 10, is equivalent to

$$a_s > \frac{A}{1 + A}(N_w^* - N_s^*). \tag{15}$$

Condition 14 requires that the imprecision of inflation control by the strong policymaker (measured by a_s) is larger than the difference in the equilibrium strategies of the two types in equation 10. When this is the case, the weak policymaker has no incentive to change the planned inflation rate in a way that would eliminate the full coverage of the strong policymaker's distribution by the weak policymaker's distribution.

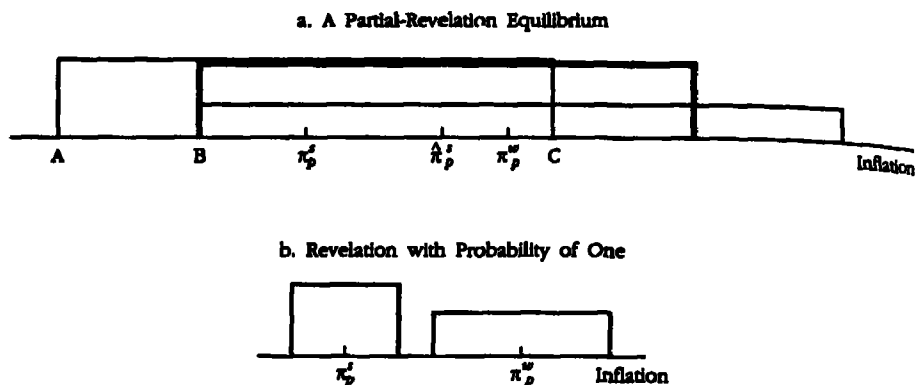
The inflation rate the weak policymaker plans to generate is greater than the rate the strong policymaker plans to generate for all periods. The weak policymaker obviously does not have an incentive to reduce the planned inflation rate, because doing so would not reduce the probability of revelation but would increase the weak policymaker's expected costs in period t. The weak policymaker also does not have an incentive to increase the inflation rate it plans to generate above its value from equation 10, because doing so would raise current expected costs and may increase the probability of revelation.[1] Equation 15 restates this condition in terms of the fundamental parameters of the model and clarifies that the condition is independent of the time index, t. The condition basically requires that the imprecision of inflation control by the strong policymaker be large compared with the difference between the desired employment objectives of the two policymakers.

Partial- versus Full-Revelation Equilibria

Unlike the weak policymaker, who has an incentive, other things being equal, to reduce the probability of revelation, the strong policymaker has an incentive to increase it. If revealed as such, the strong policymaker will reap the benefits of a good reputation during the entire future. But to increase the probability of the public's being able to separate the strong policymaker from the weak one, the strong policymaker's planned rate of inflation must be lowered. Thus the range of inflation rates that could be produced only by the strong policymaker would be widened.

If, given the weak policymaker's planned rate of inflation, the strong policymaker chooses a planned rate of inflation denoted by $\hat{\pi}_p^s$ in figure 2, the distribution of both policymakers' planned inflation rates would fully overlap, thus making the probability of sharp separation zero. But, by reducing the planned rate of inflation to π_p^s, the strong policymaker can create a positive probability

1. The implicit assumption underlying this statement is that both policymakers believe that off-equilibrium observations of inflation do not induce updating in the reputation parameter, α, and in expectations. The game theory literature refers to assumptions regarding the beliefs of players about off-equilibrium situations as "conjectures" and to the type of conjecture used here as a "passive conjecture" (Rubinstein 1985).

Figure 2. *Partial- and Full-Revelation Equilibria*

of sharp separation. In the figure this probability is measured by the ratio of linear segments AB/BC. In the extreme case, the strong policymaker can reduce the planned rate of inflation enough to create a complete separation (with a probability of one) between the two types. Figure 2b illustrates such a situation. This corresponds to the notion of a separating equilibrium in signalling theory as illustrated, for example, by the work of Vickers (1986).

When Is Gradualism the Optimal Strategy for a Strong Policymaker?

Returning to the case of strictly gradual stabilization, as in figure 1, in order for the strategy of the strong policymaker in equation 10 to be optimal, it must dominate all strategies that yield positive probabilities of sharp separation. This occurs when the current cost of deviating from the strategy in equation 10 to a strategy that opens a "window of separation" is larger than the expected present discounted value of the benefits of separation. The present cost arises because, by lowering current inflation, the strong policymaker deviates from the optimal strategy of balancing the costs of low employment and high inflation in the current period. Sharp separation yields future benefits, however, because once he or she is recognized as strong, the policymaker enjoys the benefits of higher employment and lower inflation levels associated with a perfect reputation (α = 1). The present costs of partial or full separation are more likely to be higher than its future benefits, the larger the difference between the policymakers' imprecision of inflation control ($a_w - a_s$) is compared with the difference in their desired levels of employment ($N_w^* - N_s^*$) and the lower the discount factor (β) is. In addition, it is likely that no separation will be attempted when the initial reputation is relatively high. The precise condition underlying this statement is presented in part 1 of the appendix in Cukierman and Liviatan (1991b).

The policymaker's rate of time preference affects the likelihood that the current costs of separation are larger than its future benefits. The less the strong

policymaker cares about the future, the larger the importance attributed to the current costs of separation. The role played by the relative sizes of the difference between the imprecision of inflation control by the weak and the strong policymakers ($a_w - a_s$) and the difference between the weak and strong policymakers' desired levels of employment ($N_w^* - N_s^*$) can be understood intuitively as follows. The larger the difference in imprecision of inflation control compared with the difference in the desired level of employment, the larger the divergence between the strong policymaker's optimal strategy for the current period and the strategy necessary to produce a positive probability of separation. Hence, when the difference in imprecision of inflation control is large in relation to the difference in the desired level of employment, the current costs of separation are more likely to be prohibitively high. Finally, when reputation is high to start with, the marginal future benefits of full revelation are small and therefore not worth the current costs of revelation.

IV. Features of Optimal Gradual Stabilizations and Applications

Even under the relatively favorable conditions in which a strong policymaker gets into office at the beginning of the stabilization process and remains there forever, it may be optimal to stabilize gradually. Speaking somewhat loosely, gradual stabilizations are optimal when there is a lot of noise in the control of inflation and when the policymaker, even if strong, has a high rate of time preference.

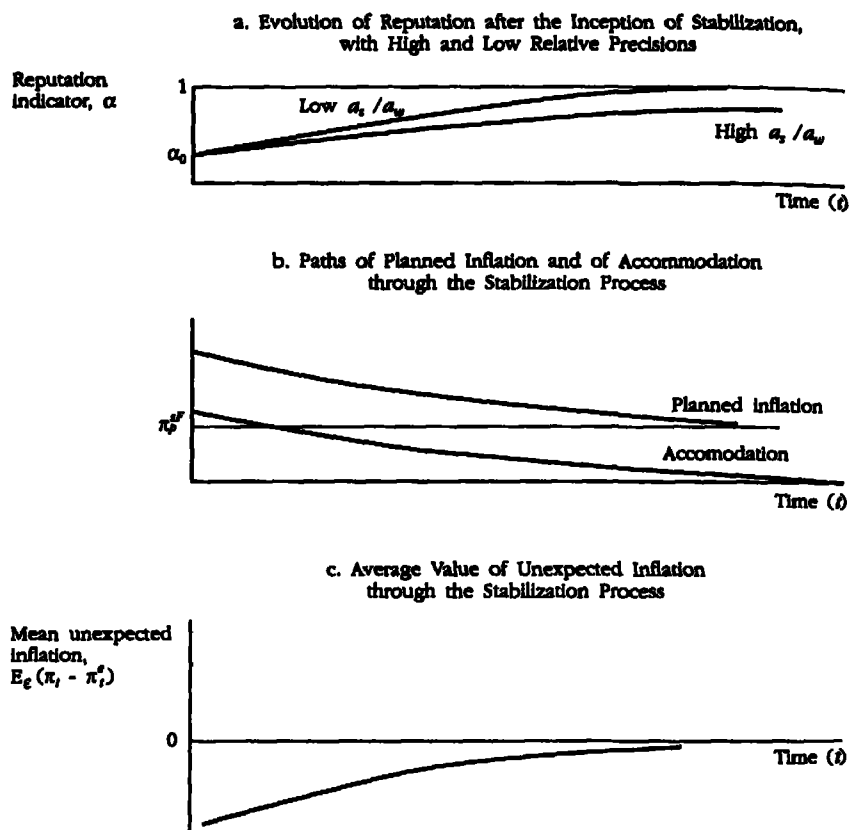
Under these circumstances the reputation of the strong policymaker as a stabilizer rises gradually (see equation 9). The speed at which it rises depends on the precision of inflation control of the strong policymaker (a_s) in relation to that of the weak policymaker (a_w). This relative precision is conveniently measured by the ratio a_s/a_w. The lower this ratio, the higher the relative precision of the stabilizing policymaker (that is, the strong one) and the faster the rate of growth of reputation. In the limit, when the ratio approaches 1, the process of reputation building may take forever. At the other extreme, when the ratio approaches 0, reputation rises extremely rapidly. Figure 3a illustrates two paths of reputation building starting from a common initial reputation α_0. In both paths, the process of reputation building is gradual, but it is faster when the ratio is lower.

From equations 11 and 12, when a firm reputation is finally established, the average rate of inflation stabilizes at the level $\pi_p^{sF} = Ad_s$, which may be a reasonably low level, provided the difference between the strong policymaker's desired level of employment and the natural level of employment is sufficiently small. But during the period of reputation buildup even a serious stabilizer (that is, the strong policymaker) partially compromises on its perfect information inflation by accommodating some of the public's suspicions. The extent of accommodation can be measured as the difference between the decisions of a serious stabilizer with and without perfect information. This measure of accom-

Figure 3. *Reputation, Accomodation, and Unexpected Inflation during Stabilization*

a. Evolution of Reputation after the Inception of Stabilization, with High and Low Relative Precisions

b. Paths of Planned Inflation and of Accommodation through the Stabilization Process

c. Average Value of Unexpected Inflation through the Stabilization Process

modation is $\pi_{pt}^s - \pi_p^{sF} = [A^2/(1 + A)]\,(d_w - d_s)(1 - \alpha_t)$, where π_p^{sF} is the strong policymaker's planned inflation rate under full information.

Thus, the larger the divergence in the employment targets of a weak and a strong policymaker and the lower the current level of reputation, the larger the extent of accommodation. As the reputation of the serious policymaker rises, the extent of accommodation diminishes monotonically and tends to disappear altogether after a sufficiently large number of periods. Concurrently, average planned inflation decreases monotonically and finally converges toward the full-information low inflation rate (π_p^{sF} in figure 3b). Figure 3b shows the paths of planned inflation and accommodation during the stabilization process. The speed at which inflation and accommodation diminish is determined by the rate of increase in reputation, which is, in turn, faster the larger the relative precision of inflation control by the strong policymaker.

Equation 13 and the fact that the difference between the desired and the natural level of employment is greater for the weak than for the strong policymaker imply that inflationary expectations also decrease monotonically during

the stabilization process and finally converge toward the low inflation rate that obtains in full information. The speed at which inflationary expectations decrease is again directly related to the relative precision of inflation control by the serious stabilizer. In spite of the fact that stabilization is gradual, the serious policymaker has to accept a prolonged period of employment below normal because the average rate of inflation is consistently lower than expected. From equations 2, 11, 12, and 13, $\pi_t - \pi_t^e = -[A/(1+A)](d_w - d_s)(1-\alpha_t) + \epsilon_t^s$.

Because the expected value of ϵ_t^s is zero, the average discrepancy between actual and expected inflation is negative on average. It is large initially and decreases monotonically toward zero as the reputation of the policymaker improves. Figure 3c depicts the path of the average discrepancy between actual and expected inflation. The tighter the relative control of the strong policymaker over inflation, the faster the rate at which unexpected inflation converges to zero. Hence, a serious stabilizer who has a tighter relative control of inflation stabilizes with a shorter and less severe reduction in economic activity. Conversely, when the policymaker's relative control is known to be loose, the recession induced by stabilization is longer and deeper.

An additional implication of the analysis concerns the dependence of the policies of both policymaker types on the evolution of reputation. From equations 11 and 12, it can be seen that, as reputation rises (as α_t goes up), the equilibrium strategies of both types gradually shift to lower rates of inflation. The intuition is that, as the reputation indicator goes up, expectations get nearer to the equilibrium strategy of the strong policymaker and further away from that of the weak policymaker. Other things being equal, this situation reduces inflationary expectations. The fact that the policies of both policymakers are positively related to expectations tends to reduce the planned rates of inflation of both types. Essentially, the increase in reputation and the associated reduction in inflationary expectations pulls both types toward lower rates of inflation.

The possibility that gradualism may be the optimal strategy sheds new light on several recent stabilization efforts in various countries. At the end of the 1970s and through the first half of the 1980s the British government under Margaret Thatcher implemented a gradual stabilization program. During this time Britain went through a prolonged period of unemployment accompanied by increasing real wages. Sargent (1986b, p. 150) criticized this gradualist approach on the grounds that it "invites speculation about future reversals, or U-turns, in policy." Sargent is right in the sense that our strong policymaker could produce a probability of one of separating himself from the weak policymaker by deflating at a sufficiently low rate. As demonstrated in the previous section, however, this course of action is not necessarily optimal. For appropriate configurations of parameters like highly imprecise control of inflation and a reasonably high initial reputation, the best stabilization strategy is gradualism.

A somewhat similar argument can be made about the disinflation process in Chile in 1974–77, where the rate of monetary growth fell rather gradually (in 1977 annual monetary growth was still above 100 percent). Contrary to a claim by Harberger (1981, 1982), a gradual reduction in monetary growth may be

consistent with a monetary crunch. Applying the principles of our model, one could argue that, under imperfect credibility, even the gradual reduction in the rate of monetary growth set by the policymaker may systematically exceed the public's expectations of this reduction. This will result in a situation of tight money. Indeed Edwards and Edwards (1987) and Corbo and Solimano (1989) conclude that monetary policy in Chile in the early stage of stabilization was contractionary.

Ireland during the 1980s also partially fits the pattern of gradualism. After 1982, both fiscal and monetary policies in Ireland became substantially more restrictive. There ensued a gradual reduction in inflation and a prolonged increase in the rate of unemployment (Dornbusch 1989). However, it is likely that part of the increase in unemployment, particularly during the second part of the 1980s, can be attributed to factors other than those modeled in this article (Blanchard and Summers 1986).

Our model provides a framework for analyzing stabilization policies in settings of moderate or high inflation, in which the main motive for inflation is high economic activity. But it does not seem to be appropriate for stabilization of hyperinflation, as experienced by Germany or Austria after World War I. Hyperinflation is a result of an unsustainable fiscal situation in which the policymaker inflates mostly or only to finance the government's budget; historically, in all major hyperinflations the revenue motive was paramount. In such cases the structure of nominal contracts becomes so condensed that the short-run tradeoff all but disappears (a in equation 4 tends to zero), and any differences in emphasis on economic activity between policymakers become inconsequential.

The credibility of stabilization relates, in this case, to the ability of the policymaker to finance expenditures in a sustainable manner (including the use of an inflation tax). This means that the issue is the credibility of the government's *solvency*. In such cases, credibility can be established (as in Sargent 1986a) more swiftly for two reasons. First, any potential differences between policymakers with regard to employment are relatively unimportant because of a shrunken Phillips tradeoff. Second, balancing of the budget (usually with foreign assistance) sends a clear signal that the major motive for inflation has been eliminated. This differs from the setting of our model, which deals with high, but not hyperinflationary, conditions, in which the credibility issue is not related to solvency, but rather to the inflation-unemployment tradeoff. In the latter setting, the gradualist solution becomes more relevant; budget balance alone does not establish quick credibility, because financing governmental expenditures is not the main motive for inflation.

V. The Effect of Foreign Aid on the Credibility of Stabilization

The path of inflation can be affected by external intervention in the form of foreign aid. An increase in the country's resources as a result of foreign aid may reduce the incentive of any type of policymaker to use surprise inflation to

increase employment and output. For example, foreign aid, by increasing the availability of raw materials and physical capital, may increase the natural levels of employment and thereby reduce the difference between the desired and the natural level of employment. Such a reduction reduces the equilibrium rate of inflation under both types of policymakers.

Foreign aid may, however, be tied more directly to the performance of the stabilization program. For example the foreign aid for any given year may be made conditional on inflation being below some target level, say π^*. If π^* is below $\pi_p^s + a_s$ in figure 1, it is clear that each policymaker will have an incentive to reduce the planned inflation rate, thus leading to a lower inflation path. The presumption is that both policymakers are better off accepting the aid and the associated conditions than rejecting this package. However, even if π^* is above $\pi_p^s + a_s$ (but below $\pi_p^w + a_w$), the strong policymaker still has an incentive to reduce the planned inflation rate, provided the following condition holds: without the constraint on the inflation rate, the weak policymaker would have inflated at a higher rate, and the public understands this fact. As a result, the public's expectation of inflation goes down compared with its expectation in the case of no conditionality. From equation 10, this decline in the expectation of inflation induces the strong policymaker to lower the planned inflation rate even further. The reason is that the public understands that tying the aid to performance would constrain the behavior of a weak policymaker.

Foreign aid conditionality will, when it is binding, motivate the weak policymaker *more* than the strong policymaker to reduce the planned rate of inflation. Consequently, the rate of inflation the weak policymaker plans to generate will move closer to the rate the strong policymaker plans to generate, and the motivation of the strong policymaker to separate will therefore diminish. Thus foreign aid conditionality will tend to make the gradualist solution more likely. However, the level of inflation with foreign aid conditionality will be lower on average.

In practice, the foregoing form of conditionality is problematic because the policymaker may be induced to produce an artificially low inflation rate by means of price controls or by similar methods (such as setting an unrealistically low exchange rate or artificially reducing public sector prices). It will therefore be advisable to formulate a more robust kind of conditionality, one that makes explicit reference not only to inflation but also to fundamentals such as the size and composition of public sector expenditures and revenues, and the stock of money or domestic credit.

VI. Conclusions

Imperfect control of inflation fundamentally alters the dynamics of inflation, reputation, expectations, and economic activity during stabilization (see also, Cripps 1991). When the difference in the ability to control inflation of strong and weak policymakers is large, unexpected inflation may be persistently nega-

tive for a while, causing reduced economic activity and giving the indication that credibility is low. But if the policymaker persists with stabilization, this pattern gradually disappears. Imperfect control also leads to a generalization of the concept of separation because it creates situations in which the equilibrium policies of different types of policymakers diverge without necessarily inducing clear-cut separation. Imperfect control of inflation is the result of factors that are related to the structure of the economy and of policymaking institutions. In particular, it is likely that the lower the degree of independence of the central bank, the lower the precision of inflation control (Cukierman 1992, chapter 18).

When the fundamental cause of inflation inertia is imperfect information about the objectives of policymakers, there is no gradualism without pain. This contrasts with the role of gradualism when the basic reason for inertia is backward-looking nominal contracts. It is possible to devise, in such cases, patterns of gradual disinflation that eliminate the employment costs of stabilization.

Appendix. Derivation of Equilibrium Strategies and Expectations under Gradualism (Equations 10 to 13)

To simultaneously solve for the equilibrium strategies and for expectations, we use the method of undetermined coefficients. In particular, we postulate that the equilibrium strategies of the two types of policymakers can be represented as the following two functions of d_i, $i = w,s$, and of expectations

$$\pi^s_{pt} = k_d d_s + k_e \pi^e_t; \; \pi^w_{pt} = r_d d_w + r_e \pi^e_t \tag{A-1}$$

where k_d, k_e, r_d, and r_e are unknown coefficients to be determined. It can be shown that only the current expectation belongs in the solution and that, given linearity and decomposability, the solution is therefore unique. There also are history-dependent trigger strategies, but we rule them out because of their limitations and lack of descriptive realism. Discussions of the coordination problem and other problems of trigger strategies appear in Rogoff (1987, 1989) and Cukierman (1992, chapter 11).

The public knows the decision rules in equation A-1 but is uncertain about the identity of the policymaker in office. Hence, inflationary expectations are given by

$$\pi^e_t = \alpha_t \pi^s_{pt} + (1 - \alpha_t)\pi^w_{pt} = \alpha_t[k_d d_s + k_e \pi^e_t] + (1 - \alpha_t)[r_d d_w + r_e \pi^e_t]. \tag{A-2}$$

Because the dynamic optimization problem in equation 8 reduces to a series of one-period problems, the equilibrium strategy of policymaker i at time t can be characterized by solving

$$\min E^i_{pt} z(d_i, \pi^i_{pt} + \epsilon^i_t, \pi^e_t), \qquad i = w,s \tag{A-3}$$

where E^i_{pt} denotes the information available to policymaker i when the policy-

maker picks the planned rate of inflation for the period. The superscript i attached to the expected value denotes the fact that the information sets of the two policymakers differ. Equation 10 is obtained from the first-order condition for the problem in A-3 and by using equation 5 and the fact that, when the policymaker picks π^i_{pt}, the policymaker takes π^e_t as given. Equating the coefficients of d_t and π^e_t across equations A-1 and 10 yields

$$k_d = r_d = k_e = r_e = [A/(1 + A)]. \tag{A-4}$$

Equation 13 follows by using A-4 in A-2. Equations 11 and 12 follow by letting $B = [A^2/(1 + A)]$.

References

The word "processed" describes informally reproduced works that may not be commonly available through libraries.

Andersen, T. M., and O. Risager. 1987. "The Role of Credibility for the Effects of a Change in the Exchange Rate Policy." Seminar Paper 377. IIES, University of Stockholm. Processed.

———. 1988. "Stabilization Policies, Credibility, and Interest Rate Determination in a Small Open Economy." *European Economic Review* 32: 669–79.

Backus, David, and John Driffill. 1985a. "Inflation and Reputation." *American Economic Review* 75: 530–38.

———. 1985b. "Rational Expectations and Policy Credibility Following a Change in Regime." *Review of Economic Studies* 52: 211–21.

Barro, R. J. 1986. "Reputation in a Model of Monetary Policy with Incomplete Information." *Journal of Monetary Economics* 17 (January): 3–20.

Barro, R. J., and D. B. Gordon. 1983. "A Positive Theory of Monetary Policy in a Natural Rate Model." *Journal of Political Economy* 91 (August): 589–610.

Blanchard, O., and L. Summers. 1986. "Hysteresis and the European Unemployment Problem." *NBER Macroeconomics Annual* 1:15–78.

Canzoneri, M. 1985. "Monetary Policy Games and the Role of Private Information." *American Economic Review* 75: 1056–70.

Corbo, Vittorio, and Andres Solimano. 1989. "Chile's Experience with Stabilization Revisited." The World Bank, Country Economics Department, Washington, D.C. Processed.

Cripps, M. 1991. "Learning Rational Expectations in a Policy Game." *Journal of Economic Dynamics and Control* 15: 297–15.

Cukierman, Alex. 1992. *Central Bank Strategy, Credibility, and Independence: Theory and Evidence*. Cambridge, Mass.: MIT Press.

Cukierman, Alex, and Nissan Liviatan. 1991a. "Optimal Accommodation by Strong Policymakers under Incomplete Information." *Journal of Monetary Economics* 27: 99–127.

———. 1991b. "The Dynamics of Optimal Gradual Stabilizations." Working Paper 34-91. Tel Aviv University, The Foerder Institute for Economic Research, Department of Economics. Processed.

Dornbusch, Rudiger. 1989. "Credibility, Debt, and Unemployment: Ireland's Fai Stabilization." *Economic Policy* 8 (April): 173–210.

Edwards, Sebastian, and A. C. Edwards. 1987. *Monetarism and Liberalization—1 Chilean Experiment.* Cambridge, Mass.: Ballinger.

Engle, R. F. 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of Variance of U.K. Inflation." *Econometrica* 50 (July): 987–1007.

Harberger, A. C. 1981. "Comment on Foxley." In W. R. Cline and S. Weintraub, ec *Economic Stabilization in Developing Countries.* Washington, D.C.: Brookii Institution.

———. 1982. "The Chilean Economy in the 1970's: Crisis, Stabilization, Liberalizati Reform." In K. Brunner and A. H. Meltzer, eds., Carnegie Rochester Confere Series on Public Policy, 17.

Kiguel, M. A., and Nissan Liviatan. 1988. "Inflationary Rigidities and Stabilizati Policies." WPS 4. World Bank, Country Economics Department, Washington, D Processed.

Kydland, F. E., and E. C. Prescott. 1977. "Rules Rather than Discretion: The Incon tency of Optimal Plans." *Journal of Political Economy* 85 (June) 473–591.

Persson, Torsten, and Sweder van Wijnbergen. 1987. "Signalling, Wage Controls, a Monetary Disinflation Policy." World Bank, Washington, D.C. Processed.

Rogoff, Kenneth. 1987. "Reputational Constraints on Monetary Policy." In K. Brun and A. H. Meltzer, eds., Carnegie-Rochester Conference Series on Public Policy, Spring, 141–82.

———. 1989. "Reputation, Coordination and Monetary Policy." In Robert Barro, e *Modern Business Cycle Theory.* Cambridge, Mass.: Harvard University Press.

Rubinstein, A. 1985. "Choice of Conjectures in a Bargaining Game with Incompl Information." In A. E. Roth, ed., *Game-Theoretic Models of Bargaining.* Cambri University Press.

Sargent, T. J. 1986a. "The Ends of Four Big Inflations." In T. J. Sargent, ed., *Ratio Expectations and Inflation.* New York: Harper and Row.

———. 1986b. "Stopping Moderate Inflations: The Methods of Poincare and Thatch In T. J. Sargent, ed., *Rational Expectations and Inflation.* New York: Harper a Row.

Vickers, J. 1986. "Signalling in a Model of Monetary Policy with Incomplete Inforn tion." *Oxford Economic Papers* 38: 443–55.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 3: 459-479

Uniform Commercial Policy, Illegal Trade, and the Real Exchange Rate: A Theoretical Analysis

Stephen A. O'Connell

Countries on fixed exchange rates sometimes use uniform tariff cum subsidy (UTCS) schemes as a way of achieving a real depreciation without disturbing the nominal exchange rate. A potential drawback of this policy in relation to an across-the-board devaluation is that a UTCS scheme provides incentives for illegal trade. Using an optimizing model with currency convertibility and illegal trade, I find that welfare is lower under a UTCS scheme than under a corresponding across-the-board devaluation and that in some cases the real exchange rate actually appreciates in response to an increase in the UTCS rate.

For countries on fixed exchange rates, achieving a real depreciation through macroeconomic policy alone may be highly contractionary if preexisting wage contracts or other rigidities prevent rapid reduction of domestic costs in relation to foreign costs. When competitiveness needs to be improved, but policy considerations prevent the use of the nominal exchange rate, a natural alternative is the uniform tariff cum subsidy (UTCS). By applying a constant ad valorem tariff to all imports and an equal subsidy to all exports, a UTCS achieves the same rapid adjustment in the domestic prices of traded goods that would be implied by a parity change.

Laker (1981) studied the use of uniform trade taxes and closely related fiscal proxies for devaluation in France (1957–58), Israel (1955–62), India (1963–66), and the Federal Republic of Germany (1968–69). UTCS schemes have been used more recently in a number of developing countries, including Ghana, Sudan, and Senegal. A recent case in point is provided by Côte d'Ivoire.

Côte d'Ivoire is a member of the West African Monetary Union, a group of countries whose common currency (the CFA franc) is freely convertible into

Stephen A. O'Connell is with the Department of Economics at Swarthmore College, Swarthmore, Pa. This article is a revision of World Bank PPR Working Paper 185 (see O'Connell 1989). The author wishes to acknowledge his indebtedness to the Macroeconomic Adjustment and Growth Division of the Country Economics Department at the World Bank for its hospitality; to the Council on Foreign Relations for its financial support; and to Bill English, Ravi Kanbur, Saul Lizondo, Dani Rodrik, Oren Sussman, and an anonymous referee for their helpful comments.

French francs by agreement with the French Treasury, which guarantees conve ibility by extending overdraft privileges to the Union's Central Bank (Krum 1985). Increases in domestic inflation starting in the mid-1970s, together wi the recent nominal appreciation of the French franc and significant nomin depreciations in neighboring Ghana and Nigeria, have produced real appreci tions in Côte d'Ivoire and a number of other CFA countries. This has led sever authors (Krumm 1987; Devarajan and de Melo 1987; Leiderman 1987) suggest that an optimal macroeconomic policy package would include a dev uation of the nominal exchange rate were it not for Côte d'Ivoire's respon bilities to the CFA Zone, where the exchange rate against the French franc h been fixed since 1948. The government of Côte d'Ivoire implemented a UT scheme in mid-1986 with the precise objective of achieving a rapid increase competitiveness.

Initial experience with the UTCS in Côte d'Ivoire has not been encouragin There is anecdotal evidence of widespread underinvoicing of imports, and some cases the government has been substantially late in paying the expc subsidy. The Ivoirien experience suggests a potentially important asymmet between devaluations and UTCS schemes: UTCS schemes, unlike devaluatior provide incentives for illegal trade. This article analyzes the implications of tl asymmetry for welfare and resource allocation within the framework of a coı monly used general equilibrium macroeconomic model. My findings suppc Laker's (1981) contention that illegal trade may significantly weaken the case f uniform trade taxes as proxies for devaluation.

I. UTCS Schemes and Dual Exchange Rates

In the literature about trade theory, the close formal relationship betwe UTCS schemes and devaluations was pointed out by Meade (1951) and, later, Bhagwati (1968) and Dixit and Norman (1980). Abstracting from issues implementation, the key difference between the two policies is one of coveraş A recent treatment is given by Adams and Greenwood (1985), who derive precise equivalence result between uniform trade taxes and exchange ra changes. They find that in a frictionless, competitive environment, a UT scheme covering trade in goods and services is fully equivalent to a devaluati of the commercial rate in a dual exchange rate system.

This asymmetry of coverage has sometimes been regarded as a primary virt of UTCS schemes. Keynes, for example, favored a UTCS as a way to address tl overvaluation of sterling in the early 1930s. He believed a UTCS scheme wou improve competitiveness without "the injury to the national credit and to o receipts from foreign loans fixed in terms of sterling which would ensue devaluation" (quoted in Laker 1981, p. 118). For the typical developing cou try, however, foreign assets and liabilities are denominated in foreign currenc so Keynes' argument does not apply.

UTCS schemes therefore represent a particularly simple departure from unifi

exchange rates. As Adams and Greenwood (1985) show, this departure does not imply an asymmetry between the effects of UTCS schemes and across-the-board devaluations, provided that the private sector (correctly) views the policies in question as permanent. With flexible prices and perfect capital mobility, neither a permanent devaluation of the commercial rate nor a permanent across-the-board devaluation has real effects. Both policies produce once-and-for-all offsetting adjustments in domestic prices and an immediate rebuilding of money balances through a balance of payments surplus. With perfect capital mobility and a fixed exchange rate, the rebuilding of money balances simply requires an exchange of foreign bonds for money between the private sector and the central bank (see Obstfeld 1981, 1986 and Adams and Greenwood 1985).

Asymmetries emerge in this idealized environment only when the policies are expected to be temporary. Adams and Greenwood (1985) and Frenkel and Razin (1989) show that, although a temporary across-the-board devaluation has no real effects, a temporary UTCS or devaluation of the commercial rate lowers the real interest rate facing the private sector and is therefore equivalent to a subsidy on foreign lending.

Actual experience with UTCS schemes suggests a further set of differences between UTCS schemes and across-the-board devaluations, differences arising from the close proximity of the former to traditional commercial policy (see Laker 1981). Unlike changes in the exchange rate, for example, UTCS schemes require an administrative apparatus for determining the values of traded goods. Moreover, the political and administrative aspects of policy implementation leave more room for rent seeking and discriminatory application in the UTCS case than in the devaluation case. Finally, as emphasized in this article, UTCS schemes drive a wedge between the domestic and foreign prices of traded goods; illegal trade therefore becomes a potentially important source of asymmetry between the two policies.

Although I will be focusing on the differences between UTCS schemes and across-the-board devaluations, the basic insights of that analysis carry over, in modified form, to the "equivalent" dual exchange rate policy. Thus, although dual (or multiple) exchange rates do not introduce direct incentives for simple smuggling or customs fraud, these systems provide indirect incentives for illegal trade as part of more complicated attempts to arbitrage between different exchange rates. A permanent devaluation of the commercial rate, for example, encourages the overinvoicing of exports and underinvoicing of imports, with the (purportedly unrelated) residual payment taking place at the more appreciated capital account rate. The exact equivalence between trade tax regimes and departures from unified nominal exchange rates is broken, however, because even without other distortions the microeconomics of illegal activity varies, depending on whether a given structure of relative prices is achieved through different exchange rates or trade taxes. Because arbitraging between dual exchange rates requires the completion of at least two separate illegal transactions, it would appear that the dual exchange rate alternative would typically provide

smaller incentives for illegal trade than the otherwise equivalent UTCs. This remains an open question, however, because the relationship between the two alternatives depends on the details of evasion technologies and enforcement policy.

II. An Equilibrium Framework Based on the Standard Model of a Dependent Economy

I will examine the implications of illegal trade using versions of the standard model of a dependent economy. Although the model is readily adaptable to short-run analysis, I will assume perfectly flexible wages and prices and therefore abstract from the short-run macroeconomic effects that would arise with nominal rigidities. I will also restrict the examination to one-period models, thereby dropping the distinctions between temporary and permanent policies and between across-the-board devaluations and devaluations of the commercial rate. And because the existence of illegal trade does not give rise to a black market in foreign currency as long as the capital account is open and free convertibility is maintained (with "no questions asked" about the source or destination of foreign exchange obtained by private individuals), I will exploit these features of the Côte d'Ivoire case and will leave monetary and portfolio issues out altogether (see Pitt 1984, Macedo 1987, and Branson and Macedo 1989 for models in which tariffs and export taxes, combined with controls on foreign exchange or capital, give rise to a black market).

I begin by reviewing the basic structure of the two-sector dependent economy model (Dornbusch 1980). The economy produces and consumes nontraded goods (N), importables (M), and exportables (X). The country is sufficiently small in world markets that the international prices of traded goods can be treated as exogenous. Given the import tariff and export tax (or subsidy) rates, the domestic prices of traded goods are also fixed, and importables and exportables can be consolidated into a single composite "traded good." The associated price index, P_T, is a linearly homogeneous function of the domestic prices of importables and exportables; in the case in which the import tariff and export subsidy are both equal to $z \geq 0$, P_T is given by

$$(1) \qquad P_T = f(P_X, P_M) = f[E(1 + z)P_X^*, E(1 + z)P_M^*] = E(1 + z)$$

where E is the nominal exchange rate, an asterisk denotes an international price, and world prices P_X^* and P_M^* are set equal to unity. The real exchange rate, e, is defined as the domestic relative price of nontraded to traded goods: $e \equiv P_N/P_T = P_N/e(1 + z)$.

I assume that labor is perfectly mobile in the period under analysis but that all other factors of production are specific to the sectors in which they are initially located. This is the Ricardo-Viner model of trade theory (see Dixit and Norman 1980 and, for a dependent economy application, van Wijnbergen 1986). The production functions $q_N(L_N)$ and $q_T(L_T)$ have the standard properties: $q(0) = 0$,

and $q'' < 0$, with the latter property (diminishing returns to labor) by the fixed supplies of sector-specific factors. For the moment, I also that $q'(0) = \infty$ in order to ensure that even a small UTCS elicits a ng response; I generalize this to $q'(0) > 0$ in the discussion at the end of ion. Wages and prices are flexible, so that full employment prevails; mobility of labor implies that a single economywide nominal wage pre- equilibrium.

ibrium can be described in a straightforward manner using revenue and iture functions (Dixit and Norman 1980). Because all activities are competitive, labor will be allocated to maximize domestic revenue from activities. Denoting the maximized value of revenues by R and assuming total supply of labor, L, $R(P_T,P_N;L)$ maximizes $P_T q_T + P_N q_N$ subject to urce constraint $L_T + L_N \leq L$. Because R is homogeneous of degree 1 in es, deflating by the domestic price of traded goods yields real revenue, $r(1,e;L)$. The revenue function r has the property that its partial deriva- e the supply functions for traded and nontraded goods.

ing to the expenditure side, aggregate expenditure on traded and non- goods is simply $P_T c_T + P_N c_N$, where c_j $(j = T,N)$ is consumption of good ninimized value of expenditure for any social utility level U is denoted by $_N;U)$. Because Z is homogeneous of degree 1 in all prices, the expenditure n can be written in terms of traded goods as $Z/P_T = \epsilon(1,e;U)$; the partial ives of ϵ are the compensated demand functions for traded and nontraded ϵ and r are related by the budget constraint $\epsilon = r - \tau$, where τ is the real f lump-sum taxes paid by the private sector.

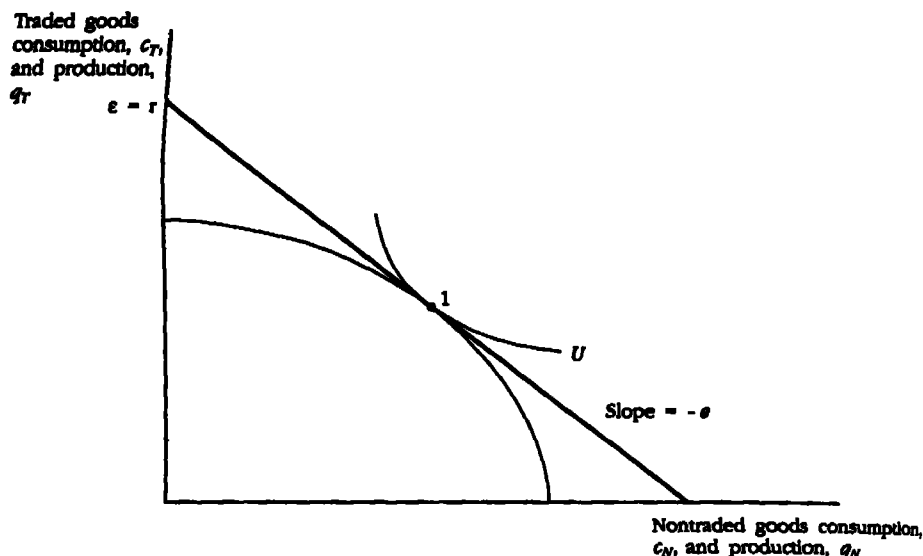
rnment finances are equally straightforward. The government receives venue from imports, pays out a subsidy to exports, and collects lump- es (which may be negative) from the private sector. Because the subsidy exports is the same as the tariff rate on imports, budget balance requires mp-sum taxes equal the official trade balance in foreign currency, b_o, ied by the UTCS rate:

$$\tau = zEb_o/P_T = [z/(1 + z)]b_o.$$

Equilibrium without Illegal Trade

out illegal trade, the official trade balance equals the overall trade bal- hich is zero in this one-period model. A UTCS scheme therefore has no net ary effect, so that $\tau = 0$. It would be straightforward to incorporate nent spending on traded and nontraded goods in the amounts g_T and g_N; case, equation 2 would read $\tau = g + zEP_T^* b_o/P_T = g + [z/(1 + z)]b_o$, $g = eg_N + g_T$.

librium prevails when there is full employment $(L_T + L_N = L)$, the led goods market clears $(\epsilon_e[1,e;U] = r_e[1,e;U])$, and the private sector vernment satisfy their budget constraints $(\epsilon = r)$. Because full employ- already subsumed in the revenue function, the latter two equations fully

Figure 1. *Equilibrium without Illegal Trade*

determine the equilibrium values of welfare and the real exchange rate (the equation for a zero trade balance is implied by these two). Figure 1 shows the familiar dependent economy diagram, with equilibrium at point 1.

The key feature of the "legal trade" equilibrium is that neither the nominal exchange rate nor the UTCS rate affects the equilibrium value of welfare or the real exchange rate. This is apparent in the fact that E and z do not enter the equilibrium conditions $\epsilon_e = r_e$ and $\epsilon = r$. A nominal devaluation or a rise in the UTCS rate simply produces an equiproportional rise in wages and the prices of nontraded goods, with no change in the real exchange rate.

Incorporating Illegal Trade

Illegal trade complicates the analysis in two ways. First, because the activity uses real resources, the economy now has at least three sectors (production of traded goods, production of nontraded goods, and production of smuggling services). Second, although uniform taxes are designed to leave the domestic relative prices of traded goods unchanged, they may fail to do so with illegal trade. When this happens, importables and exportables can no longer be consolidated into a single traded good.

Section III focuses on the first of these complications, following Bhagwati and Hansen (1973) in assuming that legal and illegal trade take place independently. Under this assumption, smuggling by competitive firms does not affect the domestic prices of traded goods unless the marginal costs of smuggling are so low that legal trade is driven out completely in equilibrium. This means that when the two types of trade coexist, as is normally observed in practice (Pitt

1981; Thursby, Jensen, and Thursby 1990), one can continue to consolidate importables and exportables into a single good. The resulting three-sector smuggling model is simple and intuitively appealing.

Although the smuggling model is appropriate for certain kinds of illegal trade (such as border smuggling), there are important categories of illegal trade (such as underinvoicing of imports or overinvoicing of exports) that use legal activity as a "cover" for illegal activity. In these cases, equilibrium is characterized by what Pitt (1981) called "price disparity": the domestic price of the imported (exported) good falls below (above) the full tariff-inclusive (subsidy-inclusive) price. Section IV, therefore, uses a model of invoicing fraud to illustrate the second channel—the domestic relative price of traded goods—through which illegal trade can affect welfare and the real exchange rate.

III. Smuggling

Smuggling firms use domestic resources to bring goods past the customs authorities. Four assumptions characterize the smuggling activity. First, smuggling services are produced by a combination of labor and factors of production specific to smuggling. Second, the cost of an additional unit of smuggling, for the smuggling industry as a whole, is an increasing function of the total amount being smuggled. Third, smuggling firms behave as price takers in the domestic market for traded goods. Fourth, the marginal cost of smuggling at the industry level rises rapidly enough that legal trade is not driven out completely in equilibrium.

The first of these assumptions implies that smuggling firms must compete with domestic producers for labor, so that both traded and nontraded goods are implicitly used up in producing smuggling services. Smuggling costs are therefore not characterized as denominated solely in either the good(s) being smuggled (Bhagwati and Hansen 1973; Pitt 1981; Martin and Panagariya 1984) or nontraded goods (Sheikh 1974).

Two kinds of market structure are consistent with the second and third assumptions. As in the Bhagwati-Hansen analysis, one can assume that although individual smuggling firms face constant costs per unit, the smuggling sector as a whole incurs increasing costs because of intra-industry, interfirm diseconomies of scale. The link between firm-level average costs and aggregate smuggling is the result of "congestion" on illegal trade routes; marginal cost pricing emerges as the result of price competition under conditions of constant returns at the firm level.

The Bhagwati-Hansen analysis implies zero profits for the smuggling sector in equilibrium. An alternative structure, in which smuggling firms are profitable in equilibrium, is one in which there is a finite number of smuggling firms, and each faces an increasing marginal cost of smuggling. In this case, it must be assumed that the number of firms, while finite, is large enough that marginal cost pricing is a good approximation to their actual strategic behavior. Although

marginal cost pricing is the case usually analyzed in the smuggling literature, the appropriateness of this assumption clearly depends on the underlying strategic environment. If smuggling firms compete over prices as Cournot rivals, for example, the smuggling industry will approach marginal cost pricing as the number of firms increases (Thursby, Jensen, and Thursby 1990). Tirole (1988) discusses the conditions under which Cournot behavior can be derived as the equilibrium of a two-stage game in which competitive suppliers first choose quantities (in our case, first smuggle a certain amount into the country) and then compete over prices.

The four assumptions make it possible to incorporate illegal trade by simply appending a third, "smuggling" sector (subscript S) to the standard analysis. To do this, let $q_S(L_S)$ be the total amount of smuggling, measured in terms of the composite traded good. The total amount of smuggling is the sum of importables and exportables smuggled into the country to avoid the tariff or receive the subsidy (an export subsidy provides an incentive for illegal importing of exportables for reexport or sale on the domestic market). Each unit smuggled yields a net revenue equal to the difference between the domestic and the world price of the good. Because smuggling occurs under increasing marginal costs at the industry level and legal trade is not driven out in equilibrium, the marginal source of supply of the traded good is legal trade, and the domestic price of the traded good will simply be the world price inclusive of the uniform tariff or subsidy rate. Real net revenue from a unit of smuggling is therefore given by

$$\frac{(P_T - EP_T^*)}{P_T} = \frac{z}{(1 + z)}.$$

As in the previous section, labor will be allocated under perfect competition so as to maximize net domestic revenue from the three activities. The revenue function therefore takes the form $r[1,e,z/(1 + z);L]$. On the expenditure side, because illegal trade leaves the price of domestic tradables unaffected, the expenditure function is still given by $\epsilon(1,e;U)$. The expenditure function and the revenue function are related through the private sector's budget constraint, $\epsilon = r - \tau$.

The previous section showed that a UTCS scheme that is implemented with a zero trade balance will have no net effect on government revenues without illegal trade. With illegal trade, however, this is no longer true: a UTCS scheme will involve a net transfer of resources to smugglers that must be made up by higher lump-sum taxes. To see this, first note that the requirement of balanced overall trade in foreign currency terms implies

$$x_O = m_S + m_O + x_S \tag{3}$$

where x and m denote the quantity of goods exported and imported, respectively, and the subscripts O and S denote official and smuggled quantities, respectively. Smuggled exports (x_S) are included among imports (as well as being part of official exports) because these are reexports illegally brought in from

other countries to take advantage of the subsidy. Equation 3 implies that the official trade surplus equals the total value of smuggling:

$$b_O = x_O - m_O = m_S + x_S = q_S. \tag{4}$$

Because lump-sum taxes, as before, equal the official trade surplus multiplied by the UTCS rate, equation 4 implies that $\tau = [z/(1 + z)]q_S$. Lump-sum taxes are exactly equal to net revenues from illegal trade.

Equilibrium

Because $q_S = r_3$ ($\equiv \partial r/\partial[z/(1 + z)]$), the real value of taxes can be written as $\tau = [z/(1 + z)]r_3$, and equilibrium can be characterized completely using the following two equations:

$$\epsilon(1,e;U) = r\left[1,e,\frac{z}{(1 + z)};L\right] - \frac{z}{(1 + z)}r_3 \tag{5}$$

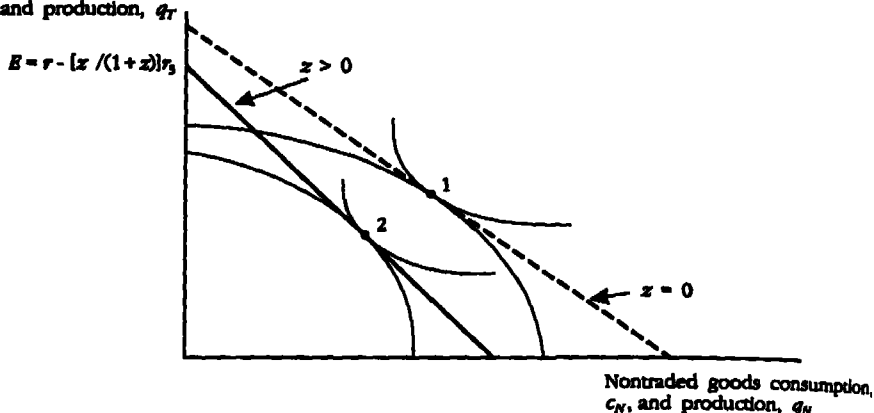
$$\epsilon_e(1;e,U) = r_e\left[1,e,\frac{z}{(1 + z)};L\right]. \tag{6}$$

Equation 5 states that, as the labor market clears, the economy is on its overall budget constraint, and the government budget constraint is satisfied; equation 6 is the market-clearing condition for nontraded goods.

Equations 5 and 6 jointly determine the real exchange rate and the social utility level as functions of the fixed total supply of labor and the UTCS rate (the third equilibrium condition, that the trade balance be zero, is implied by these two). As before, the nominal exchange rate does not appear in the equations. Any devaluation is immediately eroded by an equiproportional rise in wages and the price of nontraded goods, with no real effects. When the smuggling activity is prohibitively costly, so that $r_3 = q_s \equiv 0$, the model reduces to the standard dependent economy model, in which the real exchange rate and the social utility level are determined independently of the UTCS rate. Without smuggling, therefore, a (permanent) UTCS has no real effects.

When smuggling is present, however, changes in the UTCS rate do have real effects. From equation 5, smuggling affects overall welfare by lowering social disposable income. It is not surprising that a UTCS scheme is unambiguously welfare-worsening, because a rise in the UTCS rate increases the incentives for smuggling. As proved in the appendix and illustrated in figure 2, with a positive UTCS rate ($z > 0$ in figure 2), any increase in the UTCS rate draws resources out of productive activities and lowers welfare. Moreover, starting from a UTCS rate of zero ($z = 0$ in figure 2), any increase in the UTCS rate that is not vanishingly small will lower welfare. This is not true for across-the-board devaluations (which leave welfare unchanged); thus there is a well-defined sense in which illegal trade tends to render uniform trade taxes inferior, or "second-best," to the policy they are intended to proxy.

A related question is whether smuggling itself worsens welfare with uniform

Figure 2. *Equilibrium with Smuggling (Real Appreciation Case)*

Note: A very small increase in z starting at $z = 0$ does not affect welfare or the real exchange rate to a first approximation.

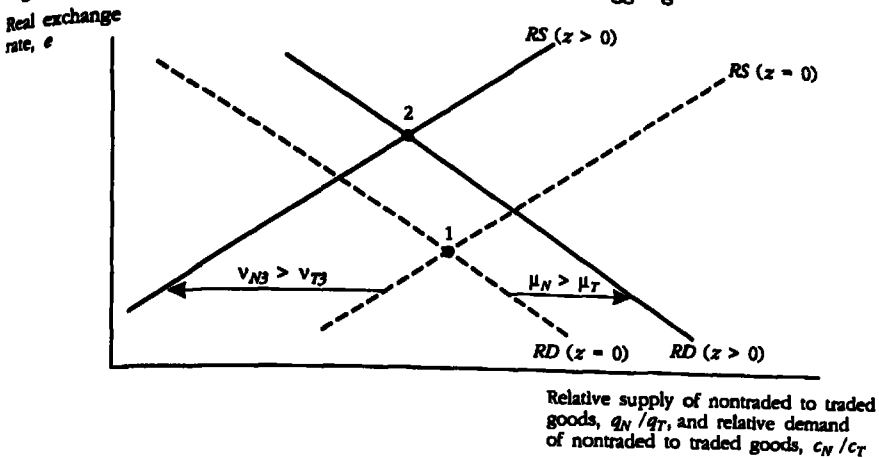
trade taxes: for a given UTCS rate, is welfare higher or lower with smuggling? The answer is again unambiguous: smuggling lowers welfare (as in Bhagwati and Hansen 1973, when legal trade is not fully displaced). The reason is that the UTCS scheme is itself not a distortion, so that the loss of productive resources as a result of smuggling occurs in an undistorted economy. If the focus had been, instead, on nonuniform commercial policies (such as import tariffs alone), the possibility of welfare gains might have emerged, because in this case the costs of illegal activity would have been offset to some degree by an amelioration of the policy distortion (Sheikh 1974; Pitt 1981; Deardorff and Stolper 1990).

The appendix shows that starting from a positive UTCS rate, the real exchange rate appreciates (rises) or depreciates (falls) according to the following criterion:

$$\frac{de}{dz} \begin{matrix} > \\ = \\ < \end{matrix} 0 \quad \text{as} \quad \frac{\nu_{N3}}{\nu_{T3}} \begin{matrix} > \\ = \\ < \end{matrix} \frac{\mu_N}{\mu_T} \tag{7}$$

where μ_j is the income elasticity of demand for good j and ν_{j3} is the cross elasticity of supply in sector j with respect to $[z/(1 + z)]$ (as with welfare, a vanishingly small UTCS rate that is introduced starting from $z = 0$ will leave the real exchange rate unchanged).

A change in the UTCS rate therefore affects both supply and demand, as illustrated in figure 3. A rise in the UTCS rate drives up the economywide wage and draws labor into the smuggling activity; the effect on the relative supply of nontraded to traded goods, q_N/q_T, depends on the relative cross elasticities of supply of these goods with respect to a rise in $z/(1 + z)$. The relative supply curve (RS in figure 3) shifts to the left (toward traded goods) if the supply

Figure 3. *Real Exchange Rate Determination with Smuggling*

Note: A very small increase in z starting at $z = 0$ does not affect the real exchange rate to a first approximation.

response is higher in nontraded goods and to the right if the response is higher in traded goods (equivalently, relative supply shifts toward the sector with a smaller elasticity of supply with respect to its own product wage, measured at the original real exchange rate). Relative supply is unchanged if $\nu_{N3}/\nu_{T3} = 1$. On the demand side, the movement of labor into smuggling produces a fall in real disposable income. The effect of this on the relative demand curve (*RD* in figure 3) depends on relative income elasticities. The relative demand curve shifts to the right if $\mu_N/\mu_T < 1$ because in this case a fall in income produces a shift in demand toward nontraded goods; the relative demand curve shifts to the left if nontraded goods have the higher income elasticity. With homothetic preferences ($\mu_N = \mu_T$), there is no effect on relative demand.

Equation 7 suggests that with illegal trade a UTCS scheme designed to achieve a real depreciation may have the opposite effect. Starting at point 1 in figure 3, where the UTCS rate is zero, a real appreciation occurs (point 2 in figure 3) if the income elasticity of demand for nontraded goods is relatively low or if the cross elasticity of supply of nontradables with respect to the UTCS rate is relatively high in nontraded goods compared with traded goods.

Income Distribution Effects

Resources devoted to smuggling end up engineering a transfer of income from taxpayers to smugglers that is mediated through the government budget. Assuming that taxes are lump-sum, the fact that these redistributions take place through the government budget has no effect on welfare or the real exchange rate. In practice, however, governments do not have access to nondistortionary

tax instruments. The reduction in public sector revenues therefore requires son combination of increases in other distortionary taxes and cuts in public sect expenditures.

With respect to welfare, as long as the shadow value of increased governme revenue is positive, the increased budgetary stringency associated with a UT scheme lowers welfare. Because these budgetary effects do not occur under devaluation, this situation further undermines the relative efficacy of unifor taxes. Effects on the real exchange rate depend on where in the budget, ar when, the required adjustment takes place. Suppose, for example, that trac taxes are the only tax instrument available to the government. The governmei budget constraint would then imply that either current government expenditu or (in a dynamic setting) future trade tax rates or expenditure must adjust to tl fall in current trade tax receipts. If current government expenditure bears tl burden, there will be upward or downward pressure on the real exchange rat depending on whether the private sector's marginal propensity to consume no traded goods is above or below that of the public sector.

Two Remarks on UTCS Schemes and Illegal Trade

The following two remarks elaborate on the comparison between the effec of UTCS schemes and devaluation.

Remark 1: High uniform tax rates and immobile labor. Although this artic has emphasized what appears to be an inexorable link between uniform trac taxes and illegal trade, in practice this link may only come into play for suff ciently high uniform tax rates. In particular, I have assumed that there are r fixed costs to initiating the smuggling activity and that the marginal product labor in smuggling is infinite at $L_S = 0$. These assumptions imply that any fini UTCS rate, regardless of how small, will move labor into smuggling. With fixe costs, however, the UTCS rate would have to reach some critical minimum lev ($\bar{z} > 0$) before there would be any smuggling response. The same would hold the marginal product of labor were finite at $L_S = 0$. In either case, the secon best nature of UTCS schemes as proxies for devaluation would emerge only for sufficiently large UTCS rate.

There would also be no real effects from a UTCS scheme in the short run labor were immobile. With immobile labor, nominal wages in all sectors wou adjust in the short run to maintain real product wages at their original level and the price of nontraded goods would rise in direct proportion to the increa in traded goods. Real effects would begin to emerge only as labor began to mo between sectors in response to real wage differentials.

Remark 2: Temporarily sticky wages and prices. Although the asymmetr between UTCS schemes and devaluations is a fundamental feature of medium- t long-run equilibrium, the analysis is more complicated if wages and prices a temporarily sticky. In this case, changes in the exchange rate can alter th relative return to legal and illegal activities in the short run, provided that th

costs of smuggling are not denominated completely in traded goods. For example, suppose that the price of nontraded goods is flexible, so that the market for nontraded goods always clears, but that unemployment arises from a sticky economywide nominal wage (that is, the economy is on the border of the Keynesian and classical unemployment regions in a disequilibrium framework). Now consider a rise in the UTCS rate or nominal exchange rate that achieves a given increase in the domestic price of traded goods, $P_T = E(1 + z)$, on impact. Because these policies lower the product wage in the traded goods sector, $w/E(1 + z)$, by the same amount, they lead to the same increase in the demand for labor there. The UTCS, however, has a greater effect on overall labor demand because it lowers the product wage in the smuggling sector, w/Ez, by a larger amount.

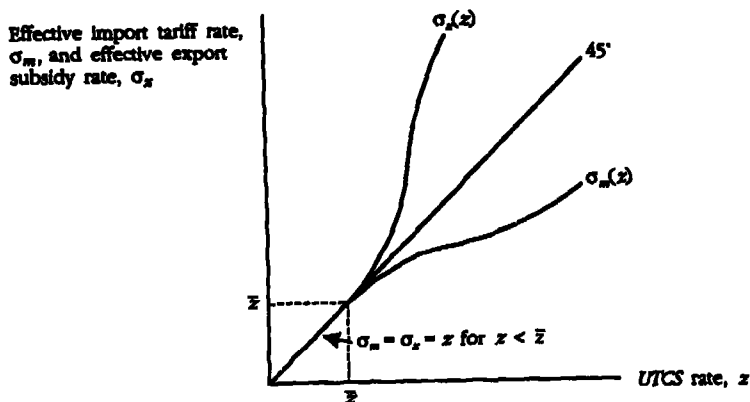
The short-run employment response is therefore larger when the change in the prices of traded goods is achieved through a UTCS than when it is achieved through a devaluation. Both policies produce an increase in employment and income, together with expenditure switching toward nontraded goods at the initial price of nontraded goods. The price of nontraded goods will therefore rise to clear that market, leading to a further increase in the demand for labor as the product wage in the nontraded goods sector falls. The final increase in employment and rise in the price of nontraded goods will be larger for the UTCS, however, given the stronger impact on employment in smuggling.

The overall welfare comparison of the two alternatives in the short run is unclear: the UTCS creates more employment, but it has more smuggling and therefore a higher resource cost. Of course, the asymmetry emphasized earlier emerges over time as wages and prices adjust to market-clearing levels.

IV. Invoicing Fraud

A channel for tariff avoidance that is important in a number of developing countries is the underinvoicing of imports. Bhagwati and Desai (1970), for example, report significant overinvoicing of exports in India in response to export incentives that were introduced along with import restrictions between 1963 and 1966 to simulate a devaluation (Laker 1981; see, also, Pitt 1981 and Bhagwati 1974). Although the use of official reference prices or specific tariffs would seem an easy solution to the problem of underinvoicing, implementation of realistic reference price systems may be very costly for nonhomogeneous imports. Côte d'Ivoire, for example, at the time the UTCS scheme was introduced, was already getting rid of specific tariffs because of their inefficiencies. In addition to these difficulties, solving the underinvoicing problem will increase the incentive for direct smuggling.

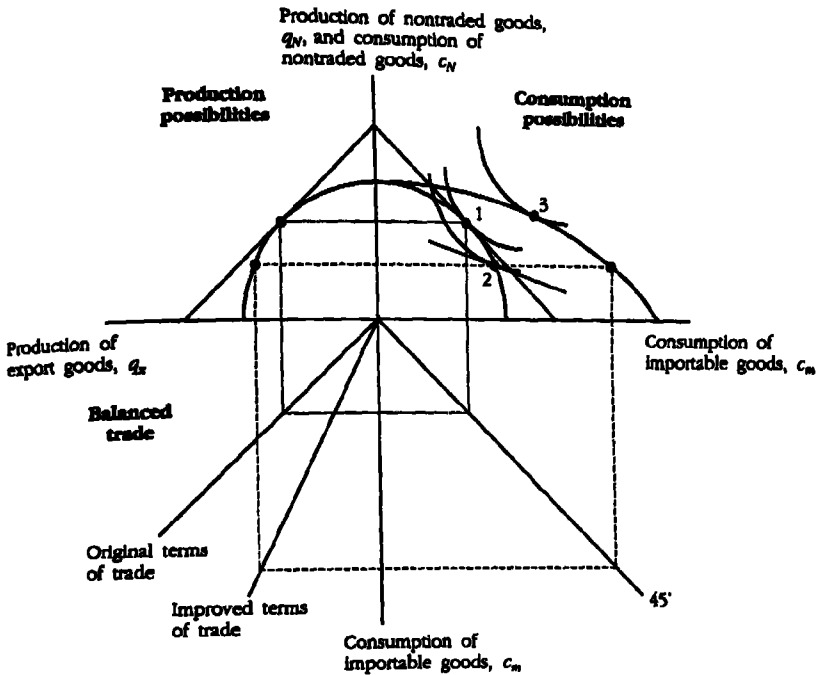
Although invoicing fraud (including overinvoicing of exports) and smuggling can both emerge in response to tax-induced divergences between the international and domestic prices of traded goods, the analysis of the previous section does not carry over directly to the underinvoicing case. There are two key

Figure 4. *Effective Tariff and Subsidy Rates with Invoicing Fraud*

Note: The diagram assumes identical penalty functions for import underinvoicing and export overinvoicing. See O'Connell (1989) for details. Penalties conditional on being caught are assumed to be a linear function of illegally appropriated revenues, plus an amount proportional to the volume of affected imports or exports. If the latter component, which acts as a fixed cost of fraudulent activity given the volume of trade, is zero, then $\bar{z} = 0$

differences in the structure of costs. First, although it was reasonable to think of smuggling as using up real resources (for example, in using inefficient transport routes), underinvoicing simply involves producing a fraudulent record of a transaction. The private cost of this activity may include bribes or penalties, but it seems appropriate to a first approximation to assume that the activity absorbs no real resources. The supply- and demand-side reallocations that were emphasized in the previous section will therefore not play a role here.

The second difference is that it is less natural in the case of invoicing fraud to think of legal and illegal trade as separate activities. In the smuggling model, the amount of smuggling could be determined independently of the extent of legal trade, because smuggling costs were independent of the amount of legal trade. No such separation is possible for customs fraud, because individuals engaged in this activity must use (the appearance of) legal trade to cover their illegal activity. This means that, in contrast to the earlier analysis, illegal trade will affect the domestic prices of traded goods. In particular, because the price of the imported good is driven below the full tariff-inclusive price and the price of the exported good is driven above the subsidy-inclusive price, the domestic price of exports in terms of imports rises as a result of an increase in the UTCS rate. This relative-price effect is the key channel through which customs fraud affects the real equilibrium.

Figures 4 and 5 bring out the main points of this analysis of invoicing fraud (a full analysis appears in O'Connell 1989). Figure 4 shows the effective import tariff and export subsidy rates, σ_m and σ_x, as functions of the UTCS rate. The effective import tariff is below the statutory tariff rate because importers underinvoice in order to lower their tariff bill, and in competition this cost saving is

Figure 5. *UTCS versus Improved Terms of Trade*

passed on to consumers in the form of a lower price. The effective export subsidy is above the statutory subsidy rate because exporters are inflating export values in order to collect a higher subsidy. Therefore

$$0 \le \sigma_m(z) \le z \text{ and } \sigma_x(z) \ge z. \tag{8}$$

Furthermore $0 < d\sigma_m/dz \le 1$ and $d\sigma_x/dz \ge 1$. Any scheme of uniform trade taxes will therefore have the same effect on the relative prices of domestic traded goods as does an export subsidy.

As shown in figure 5, this inadvertent commercial policy distorts the allocation of resources and lowers welfare. For simplicity, consider an economy in which only nontraded goods and importables are consumed and only nontraded goods and exportables produced. The upper half of the diagram shows feasible production and consumption patterns, given the economy's endowment of factors of production and the exogenous terms of trade. The balanced-trade line in the southwest quadrant takes the form $P_X^* x = P_M^* m$, where x and m are actual exports and imports, respectively. Together with the 45° line in the southeast quadrant, the balanced-trade line translates any production pattern in the northwest quadrant into the corresponding consumption pattern in the northeast quadrant.

It is apparent from the diagram that the introduction of a finite UTCS appreciates the real exchange rate for imports (the domestic relative price of nontraded goods to imports) and moves the consumption equilibrium from point 1 to point 2, where welfare is lower. On the production side, the real exchange rate for exports depreciates, and production shifts in favor of exports, which now command a higher relative price. The welfare loss occurs without any loss in real resources; welfare would be even lower in the presence of social costs of the type emphasized in the previous section.

It is instructive to compare the UTCS equilibrium with what happens when the same improvement in the domestic terms of trade occurs as a result of an exogenous movement in the external terms of trade—P^*_X/P^*_M. The difference is that an improvement in the world terms of trade represents an increase in consumption possibilities for the economy. The balanced trade locus rotates counterclockwise in the southwest quadrant of the diagram, and the consumption possibilities locus expands to reflect the economy's increased command over imports. Welfare rises to point 3. As long as both goods are normal, the real exchange rate in terms of importables appreciates; what happens to the real exchange rate in terms of exportables depends on income and substitution effects (Gavin 1988). If the income effect dominates, consumption of the nontraded good rises, and the real exchange rate in terms of exportables must depreciate. If the substitution effect dominates, consumption and production of nontraded goods fall, and the real exchange rate in terms of exportables appreciates.

The fiscal effects of a UTCS scheme with invoicing fraud are qualitatively identical to those emphasized in the smuggling case. In particular, although designed to be revenue-neutral when trade is balanced, uniform trade taxes produce a fiscal shock that must be made up for by expenditure cuts or higher taxation. As before, the presumption is that this adds to the welfare burden of the policy.

V. Conclusion

This article has examined the similarities and differences between uniform trade taxes and exchange rate changes. Without illegal trade, a UTCS scheme is equivalent in all real respects to a devaluation of the commercial rate in a dual exchange rate system.

It is not surprising that when the possibility of illegal trade is taken into account, the equivalence between uniform trade taxes and dual exchange rates is broken. More important is the fact that illegal trade undermines the attractiveness of both of these alternatives as proxies for across-the-board devaluation. These weaknesses are particularly important in developing countries, because they seem most likely to emerge when governmental administrative capabilities are thinly stretched and the required tax-cum-subsidy rate is large.

The analysis identified two channels through which illegal trade alters the

operating characteristics of UTCS schemes. The first is a resource-cost effect: when illegal trade uses up domestic resources, an increase in trade taxes drives down real income by drawing resources into smuggling, thereby producing negative supply and demand shocks in both the tradables and the nontradables sectors. With respect to the real exchange rate, it is striking that using a UTCS to raise the domestic relative price of traded goods may backfire and end up actually appreciating the real exchange rate in terms of importables. This outcome emerges if the nontradables sector has a sufficiently low income elasticity of demand or a relatively high cross elasticity of supply with respect to the UTCS rate.

The second channel is through the domestic relative price of traded goods. In the presence of some types of illegal trade, such as invoicing fraud, a rise in the UTCS rate will raise the domestic price of exports in relation to imports, appreciating the real exchange rate for imports and depreciating the real exchange rate for exports. In this case, a trade tax package that was designed to be neutral between traded goods actually ends up introducing an inadvertent net subsidy to exports.

Both the resource-cost effect and the relative-price effect produce efficiency losses for a small open economy and therefore lower welfare in relation to what would prevail under an across-the-board devaluation. These findings suggest that whatever the macroeconomic benefits of avoiding a change in the official exchange rate, illegal trade introduces a nonequivalence between UTCS schemes and across-the-board parity changes that is unambiguously unfavorable to the former.

A number of extensions of the analysis, particularly on the macroeconomic front, would help complete this account of the operating characteristics of UTCS schemes. The first is to add investment to the model and investigate the relationship between investment response, the real exchange rate, and fiscal revenues under a UTCS when the government does not have lump-sum taxes. The key issue here is that, although the tariff component of a UTCS simultaneously satisfies both the relative-price and the revenue objectives of the government, the export subsidy component brings out a conflict between these two objectives. The government may therefore have an incentive to renege on the export subsidy component of the package. If private investors anticipate this possibility, the investment response in the exported goods sector is likely to be considerably weaker than under the corresponding devaluation.

The second extension would be to introduce explicitly some wage and price stickiness so as to reinstate the real balance and relative-price effects that are the traditional channels for short-run real effects of devaluations and UTCS schemes. I emphasized in the introduction that parity changes are often defended on the grounds that they directly lower the real wage in terms of traded goods, thereby helping to avoid "transitional unemployment" associated with more gradual approaches to real depreciation. On the negative side, this short-run gain must be weighed against the other short-run, primarily contractionary, effects of par-

ity changes or UTCS schemes that appear when prices are sticky. It would help to know how, if at all, these short-run tradeoffs are affected by the illegal trade associated with the UTCS option.

Finally, although I have assumed perfect foresight and permanent changes in trade taxes or exchange rates, it is worth emphasizing that the balance of payments effects of UTCS policies depend crucially on how these policies affect expectations regarding the financial exchange rate. Laker (1981, p. 137) found that "in all four countries [surveyed in his article], the fiscal proxies were subsequently abandoned in favor of an explicit change in the exchange rate." This suggests that introducing a UTCS scheme is likely to raise subjective probabilities of devaluation of the financial rate (either alone or as part of an across-the-board devaluation). This in turn means an increase in the domestic nominal interest rate and a portfolio shift away from domestic currency and toward foreign bonds. If the interest elasticity of money demand is high enough, the overall balance of payments may well deteriorate as a result. Indeed, if there are limits to international borrowing by the central bank, it may be impossible (without other policy action) for the authorities to rule out an equilibrium in which implementation of the UTCS leads to self-fulfilling expectations of a balance of payments crisis and devaluation of the financial rate.

Appendix. Welfare and the Real Exchange Rate in the Smuggling Model

Totally differentiating equations 5 and 6 in the text, gives

$$\text{(A-1)} \quad \begin{bmatrix} de \\ dU \end{bmatrix} = \frac{1}{\Delta} \cdot \begin{bmatrix} \epsilon_{eU} & -\epsilon_U \\ r_{ee} - \epsilon_{ee} & \dfrac{z}{(1+z)} r_{e3} \end{bmatrix} \begin{bmatrix} -z(1+z)^{-1} r_{33} \\ r_{e3} \end{bmatrix} \cdot d[z/(1+z)]$$

where subscripts e, U, and 3 denote partial derivatives with respect to e, U, and $[z/(1+z)]$, and where $\Delta \equiv [z/(1+z)] r_{e3}\epsilon_{eU} + \epsilon_U(r_{ee} - \epsilon_{ee}) = r_{ee}\epsilon_{1U} - r_{e1}\epsilon_{eU} - \epsilon_U\epsilon_{ee} > 0$. Thus,

$$\text{(A-2)} \quad \frac{dU}{dz} = \frac{z}{(1+z)^3} \frac{1}{\Delta} \left([r_{e3}^2 - r_{ee} r_{33}] + \epsilon_{ee} r_{33}\right)$$

$$\text{(A-3)} \quad \frac{de}{dz} = \frac{z}{(1+z)^2} \frac{1}{\Delta} \left[\epsilon_{eU} r_{31} - \epsilon_U r_{e3}\left(1 - \frac{e\epsilon_{eU}}{\epsilon_U}\right)\right]$$

where equation A-3 uses the fact that $r_{31} + e r_{3e} + [z/(1+z)] r_{33} = 0$, by homogeneity of degree 1 of r_3.

Consider the effect on welfare first. By convexity of the revenue function, the term in square brackets in equation A-2 is nonpositive; this implies that the entire expression is negative, except at $z = 0$, where it is zero (although an infinitesimal rise in z starting at $z = 0$ produces a first-order increase in smuggling services, it does so without drawing more than an infinitesimal amount of

labor from productive sectors and therefore without affecting overall income or utility to first order). UTCS schemes therefore unambiguously worsen welfare in this model: any finite increase in the UTCS rate draws resources out of productive activities and lowers welfare.

Now consider de/dz. Under the assumption that the marginal product of labor in smuggling goes to infinity as q_S goes to zero, one can show that $r_{13} = r_{e3} = 0$ for $z = 0$. An infinitesimal change in z starting at $z = 0$ therefore has no effect on e. For $z > 0$, however, r_{e3} is strictly less than zero. To derive equation 7, note that $\epsilon = \epsilon_1 + e\epsilon_{e1}$, by homogeneity of degree 1 of ϵ. This implies $\epsilon_U = \epsilon_{1U} + e\epsilon_{Eu}$. Using this to substitute for the last term in equation A-3, gives

$$\frac{de}{dz} = \frac{1}{(1+z)^2} \frac{1}{\Delta} (\epsilon_{eU} r_{13} - \epsilon_{1U} r_{e3}). \tag{A-4}$$

Income elasticities of demand for the traded and nontraded goods are given by $\mu_T = \epsilon_{1U}\epsilon/\epsilon_U\epsilon_1$ and $\mu_N = \epsilon_{eU}\epsilon/\epsilon_U\epsilon_e$, respectively. Using these, equation A-4 implies

$$de/dz \gtreqless 0 \text{ as } \frac{r_{e3}/\epsilon_e}{r_{13}/\epsilon_1} \gtreqless \frac{\mu_N}{\mu_T}. \tag{A-5}$$

The cross elasticities of supply of nontraded and traded goods with respect to $[z/(1+z)]$ are $\nu_{N3} = r_{e3}[z/(1+z)]/r_e$ and $\nu_{T3} = r_{13}[z/(1+z)]/r_1$. Because $r_e = \epsilon_e$ by market clearing and $r_1 = \epsilon_1$ by a zero trade balance, the last expression in equation A-5 is ν_{N3}/ν_{T3}. This yields equation 7 in the text.

References

The word "processed" describes informally reproduced works that may not be commonly available through libraries.

Adams, Charles, and Jeremy Greenwood. 1985. "Dual Exchange Rate Systems and Capital Controls: An Investigation." *Journal of International Economics* 18 (1/2, February): 43–63.

Bhagwati, Jagdish. 1968. "The Theory and Practice of Commercial Policy: Departures from Unified Exchange Rates." *Special Papers in International Economics*. 8. Princeton University, Department of Economics, International Finance Section, Princeton, N.J.

———, ed. 1974. *Illegal Transactions in International Trade*. Amsterdam: North-Holland.

Bhagwati, Jagdish, and Padma Desai. 1970. *India: Planning for Industrialization—Industrialization and Trade Policies since 1951*. Oxford, U.K.: Oxford University Press for the Development Centre of the Organisation for Economic Co-operation and Development.

Bhagwati, Jagdish, and Bent Hansen. 1973. "A Theoretical Analysis of Smuggling." *Quarterly Journal of Economics* 87(2, May): 172–87.

Branson, William H., and Jorge Braga de Macedo. 1989. "Smugglers' Blues at the Central Bank: Lessons from Sudan." In Guillermo Calvo, Ronald Findlay, Pentti Kovri, and Jorge Braga de Macedo, eds., *Debt, Stabilization, and Development.* Oxford, U.K.: Basil Blackwell.

Deardorff, Alan, and Wolfgang Stolper. 1990. "Effects of Smuggling under African Conditions: A Factual, Institutional, and Analytic Discussion." *Weltwirtschaftliches Archiv* 126(1): 116–41.

Devarajan, Shantayanan, and Jaime de Melo. 1987. "Adjustment with a Fixed Exchange Rate: Cameroon, Côte d'Ivoire, and Senegal." *The World Bank Economic Review* 1 (3): 447–87.

Dixit, A. K., and Victor Norman. 1980. *Theory of International Trade.* Cambridge: Cambridge University Press.

Dornbusch, Rudiger. 1980. *Open Economy Macroeconomics.* New York: Basic Books.

Frenkel, Jacob, and Assaf Razin. 1989. "Exchange-Rate Management Viewed as Tax Policies." *European Economic Review* 33(4, April): 761–81.

Gavin, Michael. 1988. "Structural Adjustment to a Terms of Trade Disturbance: The Role of Relative Prices." *Journal of International Economics* 28(3/4, May): 217–43.

Krumm, Kathie. 1985. "Adjustment Policies in the Ivory Coast in the Framework of UMOA." CPD Discussion Paper 1985–6. World Bank, Country Policy Department, Washington, D.C. Processed.

———. 1987. "Adjustment in the Franc Zone: Focus on the Real Exchange Rate." CPD Discussion Paper 1987-7. World Bank, Country Policy Department, Washington, D.C. Processed.

Laker, John F. 1981. "Fiscal Proxies for Devaluation: A General Review." *IMF Staff Papers* 28 (1, March): 118–43.

Leiderman, Leonardo. 1987. "Price and Wage Flexibility in the Adjustment Process of Franc Zone Countries: The Ivory Coast." University of Chicago, Department of Economics. Processed.

Macedo, Jorge Braga de. 1987. "Currency Inconvertibility, Trade Taxes, and Smuggling." *Journal of Development Economics* 27 (1/2, October): 109–25.

Martin, Lawrence, and Arvind Panagariya. 1984. "Smuggling, Trade, and Price Disparity: A Crime-Theoretic Approach." *Journal of International Economics* 17 (3/4, November):201–17.

Meade, J. E. 1951. *The Balance of Payments: The Theory of International Economic Policy.* Vol. 1. London: Oxford University Press.

Obstfeld, Maurice. 1981. "Capital Mobility and Devaluation in an Optimizing Model with Rational Expectations." *American Economic Review: Papers and Proceedings* 71 (2, May): 217–21.

———. 1986. "Capital Controls, the Dual Exchange Rate, and Devaluation." *Journal of International Economics* 20 (1/2, February): 1–20.

O'Connell, Stephen A. 1989. "Uniform Trade Taxes, Devaluation, and the Real Exchange Rate: A Theoretical Analysis." PPR Working Paper 185. World Bank, Country Economics Department, Washington, D.C. Processed.

Pitt, Mark M. 1981. "Smuggling and Price Disparity." *Journal of International Economics* 11 (4, November): 447–58.

———. 1984. "Smuggling and the Black Market for Foreign Exchange." *Journal of International Economics* 16 (3/4, May): 243–57.

Sheikh, Munir A. 1974. "Smuggling, Production, and Welfare." *Journal of International Economics* 4 (4, November): 355–64.

Thursby, Marie, Richard Jensen, and Jerry Thursby. 1990. "Smuggling, Camouflaging, and Market Structure." Purdue University, Department of Economics, West Lafayette, Ind. Processed.

Tirole, Jean. 1988. *The Theory of Industrial Organization*. Cambridge, Mass.: MIT Press.

van Wijnbergen, Sweder. 1986. "Macroeconomic Aspects of the Effectiveness of Foreign Aid: On the Two-Gap Model, Home Goods Disequilibrium and Real Exchange Rate Misalignment." *Journal of International Economics* 21 (2, August): 123–36.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 3: 481-502

Maize and the Free Trade Agreement between Mexico and the United States

Santiago Levy and Sweder van Wijnbergen

Setting the price of maize in rural Mexico above the world price is inefficient and likely to have negative distributional effects because many subsistence producers, and all landless workers, are net buyers; in fact it screens out the relatively poor rather than the relatively rich. The policy objective, therefore, should be to move toward free trade. This would yield large gains in efficiency.

The Free Trade Agreement provides an ideal opportunity to pursue this objective. It will provide freer entrance into the United States for other agricultural products as well as a broad range of manufactured products. Insuring secure and sustained access for labor-intensive agricultural and manufactured products can help ease the impact on the labor market of a transition away from subsistence maize cultivation.

Maize is perhaps the single most important commodity in Mexico. In rural areas it is the main food consumed by farmers; in urban areas it is the main input into tortillas, a key component of urban workers' diets. Maize cultivation occupies between one-third and one-half of the country's arable land and employs one out of three rural workers. It is grown by a large number of small-scale producers on rain-fed land and by relatively fewer large-scale farmers on irrigated land. But because many small-scale producers, or subsistence farmers, have plots of very poor quality, maize is directly associated with rural poverty. In addition poverty in Mexico is to a large extent a rural phenomenon (Levy 1991).

Governments in Mexico have announced maize self-sufficiency as a national goal. Governments also have expressed their commitment to poor maize producers by subsidizing production. The process of land reform in Mexico gave farmers some land, but, as the extensive margin was exhausted, the quality of the land distributed diminished. Of the 43 million hectares distributed between 1958 and 1976, 91 percent were hillside, mountainous terrains; 8.4 percent were rain-fed land; and only 0.4 percent were irrigated land (Salinas 1990).

Raising the producer price of maize was one way to increase the value of the

Santiago Levy is with the Economics Department at Boston University. When this article was written, Sweder van Wijnbergen was the lead economist for Mexico at the World Bank. The authors thank Hans Binswanger for useful discussions, Hartwig Schafer for valuable assistance, and the three anonymous referees for very helpful suggestions.

asset distributed. Governments have also expressed their commitment to the poor by subsidizing maize consumption, although the system of support to consumers operates mostly in urban areas. Rural consumers obtain most of their maize at the producer price. Attempts are made to subsidize rural consumers through a network of Conasupo (the government's food marketing and distribution agency) stores, where maize is sold at a discount. But these subsidies are small and do not systematically reach rural consumers. Therefore, rather than referring to the consumer and producer prices of maize, we instead refer to the urban and rural prices of maize.

To accomplish its goals of maize self-sufficiency and support for poor maize producers, the government controls imports, and intervenes directly in marketing and distribution through Conasupo. These policies thus raise distributional issues between urban and rural areas, and within rural areas, because not all rural producers grow maize and only a subset of maize producers are net sellers. But because significant land and labor resources are allocated to this crop, maize policies also have important effects on efficiency.

The soon-to-be-negotiated Free Trade Agreement (FTA) between Mexico and the United States has placed maize at the forefront of policy debates in Mexico. Policymakers face a dilemma: continue present policies, or include maize in the FTA. This article analyzes maize pricing policies in Mexico and explicitly calculates the costs of keeping maize outside the FTA. We argue that efficiency and distributional gains can be made by liberalizing maize. For a general analysis of agricultural pricing in developing countries, see Sah and Stiglitz (1987); for analysis of individual countries, see, for example, Braverman, Hammer, and Gron (1987) for Cyprus and Newbery (1987) for the Republic of Korea.

I. A Framework for Analysis

We opt for a partial equilibrium approach mostly because of the forbidding data requirements for a full-fledged general equilibrium model. This is in the spirit of much recent work on taxation in developing countries (Newbery and Stern 1987; Newbery 1987). We show that the tools used in this approach are in fact quite flexible and easily adaptable to reflect much country-specific detail. Sah and Stiglitz (1987) use similar tools, but go quite far toward a general equilibrium analysis. Their focus on the consequences of rural-urban interaction on growth is different from ours, however.

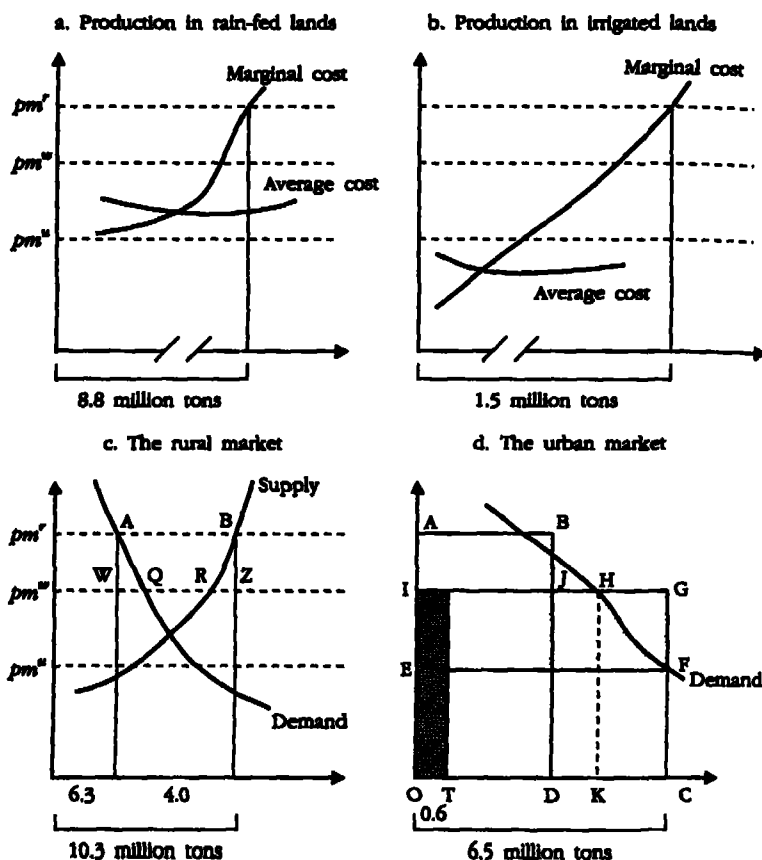
Four groups are associated with maize in Mexico. First, small-scale producers consist of farmers with three hectares or less of rain-fed land, who derive their income partly from producing maize and partly from participating in the labor market; they consume a significant share of their own production. Second, landless rural workers derive their income from wages and at times compete with subsistence producers on the supply side of the rural labor market. Third, large-scale producers mostly own irrigated land, are net buyers of labor, and

eir income from producing a wide variety of crops. Fourth, urban
rs comprise many groups with different income levels and expenditure

luate maize policies, we distinguish between income changes incurred
dual groups and aggregate welfare. Calculations of aggregate welfare
ymakers whether a particular policy needs to be changed; calculations
ncome loss or gain per group indicate which groups might need
e.

The Market for Maize

pricing policies in 1989 fixed the rural price, p_m^r, above the world
, but set the urban price, p_m^u, below the world price. In this respect
policies differ from those of many other developing countries, where
urban and rural prices of the key staple are below the world price, or
middle-income countries like Korea, where the rural price of rice ex-
urban price, but both exceed the world price (Newbery 1987). Figure 1
s the 1989 interventions in the market for maize in Mexico. Panel a
erage and marginal cost curves for maize production on rain-fed land;
loes the same for irrigated land. Average costs on rain-fed land are
an on irrigated land. In addition, the supply curve of maize on irrigated
ore elastic, which reflects the fact that irrigated land gives greater crop
producers, thus resulting in higher cross-price elasticities of supply
maize and other crops. The rural maize market is depicted in panel c.
ly curve is obtained from the horizontal addition of the marginal cost
f the two types of producers; the demand curve reflects own-
tion and other rural uses of maize (animal feed and seeds). At p_m^r the
or is a net exporter, in the amount AB. Increasing p_m^r above p_m^W gives
ubsistence producers but gives larger inframarginal rents to large-scale
s on irrigated land. Panel d shows that at p_m^u urban demand for maize is
s is met by imports from rural areas in the amount AB (purchased by
o at the price p_m^r), together with imports from abroad in the amount
hased by Conasupo at the price p_m^W). Since urban consumers pay p_m^u for
mount OC, this pricing-cum-import control scheme requires a subsidy
the shaded areas of panel d, or the sum of rectangles AIJB and EFGI.
naize producers are subsidized, rural maize consumers are taxed, and
ize consumers are subsidized. Rectangle EFGI is the subsidy received
consumers; rectangle AIJB is the subsidy received by rural producers.
fects on imports and the fiscal balance of equating the rural and urban
maize to the world price are as follows. At p_m^W production by both types
cers declines, together with land rents. Rural consumption increases,
luces net maize exports to the urban sector from AB to QR. But at p_m^W
nsumption falls from OC to OK. This partly reduces the increase in
associated with lower domestic supply. In fact, it is conceivable that

Figure 1. *The Market for Maize in Mexico*

imports could fall (if, say, rural maize demand is inelastic and urban consumption represents a large amount of total consumption). Eliminating subsidies generates fiscal savings measured by the shaded area of panel d.

Real Income Effects: Direct Price Effects

Consider a simple set-up where there are only three tradable products: maize, vegetables (a proxy for all other agricultural products), and manufactures (denoted by subscripts m, v, and q, respectively). Increasing the number of goods leaves the argument intact as long as their prices are not affected by changes in the price of maize, as is the case for tradable goods. Introducing nontradables complicates the argument because we need to incorporate the effect of changing maize prices on the price of the nontradable. But in Mexico maize accounts for less than 1 percent of gross national product, so even large changes in maize prices are unlikely to have an important effect on the price of nontradables.

To sharpen the argument, we also assume that rain-fed lands can produce only maize, while irrigated lands can also produce vegetables, and that all rain-fed lands are owned by subsistence producers and all irrigated lands are owned by large-scale producers. Let p_v and p_q be the prices of vegetables and manufactures, respectively, and let manufactures be the numeraire, so that $p_q = 1$. Prices of vegetables and manufactures are assumed to be the same in urban and rural areas. Abstracting from transport costs, this is a fair representation of Mexico, because there is no agency like Conasupo that explicitly intervenes to set a wedge between the two. Finally, let w^r be the rural wage rate. Consider now the income effects on each of the four groups mentioned above of removing production and consumption subsidies to maize.

Subsistence producers. Subsistence producers must allocate their total labor, $\bar{L}_p$ (where an overbar denotes an exogenous variable), between cultivating maize on their rain-fed land, L_m^{rf}, and participating in the labor market $(\bar{L}_p - L_m^{rf})$. Writing Q_m^{rf} for the quantity of maize produced by subsistence producers on rain-fed lands, and $E(\cdot)$ for the expenditure function, the budget constraint for subsistence producers is

$$(1) \qquad E(p_m^r, p_v, 1, U) = p_m^r \cdot Q_m^{rf}(L_m^{rf}) + w^r \cdot (\bar{L}_p - L_m^{rf})$$

where U is utility level. We ignore leisure, given the valuation difficulties in rural environments; see Braverman, Hammer, and Ahn (1987) for further discussion. Differentiating equation 1 with respect to the rural maize price, p_m^r, yields

$$(2) \quad E_{p_m^r} \cdot dp_m^r + E_U \cdot dU = p_m^r \cdot (dQ_m^{rf}/dp_m^r)dp_m^r + Q_m^{rf} \cdot dp_m^r + (\bar{L}_p \cdot dw^r/dp_m^r - w^r \cdot dL_m^{rf}/dp_m^r - L_m^{rf} \cdot dw^r/dp_m^r)dp_m^r$$

where we use the fact that $dp_v/dp_m^r = 0$. Noting that $E_{p_m^r}$ is the compensated demand for maize, C_m, that at an optimal labor allocation $p_m^r \cdot \partial Q_m^{rf}/\partial L_m^{rf} = w^r$, and that $dQ_m^{rf}/dp_m^r = (\partial Q_m^{rf}/\partial L_m^{rf}) \cdot dL_m^{rf}/dp_m^r$, we can rearrange equation 2 to yield

$$(3) \qquad E_U \cdot dU = (Q_m^{rf} - C_m) \cdot dp_m^r + (\bar{L}_p - L_m^{rf})(dw^r/dp_m^r)dp_m^r.$$

E_U is the inverse of the marginal utility of income, so the left side of equation 3 provides a "money" measure of the change in real income associated with a small change in the rural maize price. Hence, equation 3 shows that changes in the rural maize price affect subsistence producers through two channels, which we label the direct price effect and the indirect wage rate effect (the first and second terms on the right side of equation 3, respectively). The direct price effect shows that without changes in the rural wage rate, the real income effect on subsistence producers of changes in the rural maize price depends only on their net maize position: subsistence producers who are net sellers (buyers) lose (gain) with a fall in the rural maize price. The indirect wage rate effect, however, shows that subsistence farmers gain (lose) on the amount of their marketed labor as the rural wage rate increases (decreases).

Labor allocated by subsistence producers to on-farm maize production is obtained by solving the equation $p^r_m \cdot \partial Q^{rf}_m / \partial L^{rf}_m - w^r = 0$. Let the solution be $L^{rf}_m(w^r, p^r_m)$, and note that $\partial L^{rf}_m / \partial p^r_m > 0$. Hence, reducing the rural maize price increases subsistence farmers' participation in the labor market. This increase in the rural labor supply causes a change in the rural wage rate, which in turn affects subsistence farmers by changing the income they receive on their marketed labor.

The term dw^r / dp^r_m captures this indirect effect. The direction of change in the wage rate, however, also depends on what happens to the demand for rural labor as the rural maize price falls. If $dw^r / dp^r_m > 0$ there are two effects. First, subsistence producers who are net sellers lose both because their marketed maize is worth less and because their marketed labor is worth less. Second, subsistence producers who are net buyers gain from lower maize prices because of the maize they buy, but lose from lower wages for their labor, so the impact on their real income is ambiguous. Conversely, if $dw^r / dp^r_m < 0$, subsistence producers who are net sellers face an ambiguous real income change (losing on their maize sold but gaining on their labor sold), and subsistence producers who are net buyers unambiguously gain (paying less for the maize they buy and getting more for the labor they sell).

Landless rural workers. By definition landless rural workers own no land; market all their labor, L_r; and purchase all the maize they consume. The change in their real income is given by

$$E_U \cdot dU = -C_m \cdot dp^r_m + L_r \cdot (dw^r / dp^r_m) \cdot dp^r_m. \tag{4}$$

Clearly, the direct effect of lower maize prices is beneficial to landless rural workers. However, the indirect effect may hurt them if the rural wage rate falls as a result of the increased labor market participation by subsistence producers and if such a fall is large enough to eliminate the gains associated with a lower price for the maize they consume. When p^r_m falls the wage rate measured in terms of maize increases, but if $dw^r / dp^r_m > 0$ the wage rate in terms of manufactures decreases.

Large-scale farmers. We assume that large-scale farmers derive all their income from the (irrigated) land they own. Their problems are to allocate their total land, T, between maize and vegetables (T^i_m and T_v, respectively,) and to determine how much labor to employ in each crop (L^i_m and L_v). Hence, their budget constraint is

$$E(p^r_m, p_v, 1, U) = p^r_m \cdot Q^i_m(T^i_m, L^i_m) + p_v \cdot Q_v(T_v, L_v) - w^r \cdot (L^i_m + L_v) \tag{5}$$

where Q^i_m is maize output on irrigated land, and Q_v is vegetable output. Efficient allocation of land requires $dT_v / dp^r_m = -dT^i_m / dp^r_m$, given the land constraint $T = T_v + T^i_m$. Differentiating equation 5 and using this condition yields

$$E_U \cdot dU = (Q^i_m - C_m) \cdot dp^r_m - (L_v + L^i_m) \cdot (dw^r / dp^r_m) \cdot dp^r_m. \tag{6}$$

Large-scale farmers certainly consume less maize than they produce. Hence, the direct price effect lowers their real income. The indirect wage rate effect again depends on the sign of dw^r/dp^r_m. If it is positive, large-scale producers gain because their total wage bill is less. If the wage fall is large enough, these gains can offset the loses on their marketed maize, thus implying that large-scale farmers may actually benefit from lower maize prices.

Urban consumers. The impact on the real income of urban workers is given by an expression similar to equation 4, except that the urban wage rate, w^u, and the urban price of maize are the relevant variables in this case. Since $p^u_m < p^W_m$, a full liberalization of the maize market would lower urban workers' real income. And urban workers could be affected through the indirect wage rate effect if rural employment contracts as a result of a large fall in the rural maize price and this, through migration, lowers the urban wage. Urban employers will also see their real income lowered when the urban maize price is increased through the direct price effect, although its significance is probably minimal because maize is relatively unimportant in their diets. However, to the extent that the indirect wage rate effect puts downward pressure on the urban wage rate through migration, the product wage in manufacturing falls, thus leading to an increase in employment and quasi-rents on the capital stock employed in manufacturing.

Real Income Effects: Indirect Wage Rate Effects

Lowering the rural price of maize increases participation by subsistence farmers in the rural labor market. But the resulting change in the rural wage rate also depends on the change in the demand for rural labor as the rural maize price falls and on the change in the size of the rural labor force resulting from migration to urban areas. In the appendix we discuss the relationship between urban and rural labor markets. Here we consider the isolated rural labor market described by the equilibrium condition in equation 7,

$$(L_p - L^{rf}_m) + L_r = L_v + L^i_m + L_g. \qquad (7)$$

The bracketed term on the left side of equation 7 is marketed labor by subsistence farmers, and L_r is marketed labor by landless rural workers. The right side is the demand for rural labor, made up of employment in vegetable and maize production on irrigated land and a term (L_g) that represents an exogenous component of rural labor demand (associated, say, with government infrastructure projects).

When the rural maize price falls, marketed labor supply of subsistence farmers increases, labor demand in irrigated maize falls, and labor demand in vegetables increases. Thus, the pressures on the rural wage rate hinge on:

$$\left|\frac{\partial L^{rf}_m}{\partial p^r_m} + \frac{\partial L^i_m}{\partial p^r_m}\right| \lesseqgtr \frac{\partial L_v}{\partial p^r_m} \qquad (8)$$

that is, on whether the additional employment created in vegetable production

can absorb the employment displaced from maize cultivation. The appendix derives conditions that determine the direction of the inequality. Here we discuss the case where there is a net release of labor (the left side of expression 8 exceeds the right side); if this is not the case, the analysis needs to be modified accordingly.

Equilibrium in the rural labor market can be restored through different mechanisms. First, ignoring migration, the rural wage rate would fall, with employment increasing in both maize and vegetable production until excess labor disappears. Second, still ignoring migration, released labor could be absorbed through direct interventions like public work programs (increase L_g). Third, equilibrium can be restored through migration. If the only policy change is a decrease in the rural maize price, the labor released from maize would reduce the rural wage rate and widen the rural-urban wage differential. This would induce rural workers to migrate to urban areas, which would mitigate the decline in the rural wage rate but would lower the urban wage rate. A lower urban wage rate, in turn, increases employment and the marginal product of capital in manufacturing. A fourth possibility arises if the decrease in the rural maize price is accompanied by an increase in the urban maize price (as urban subsidies are eliminated). In this situation migration incentives are reduced because there are offsetting changes. The outcome in terms of migration and wage rate changes depends on the magnitude of the different effects. Clearly, however, policymakers can influence migration flows (and hence the changes in the rural and urban wage rates) through both direct interventions, which hire labor for rural public work programs, and the size of programs like the *tortibonos* in Mexico, which is targeted on the urban poor and operates through coupons.

Aggregate Social Welfare

In the rural sector we treat the difference between the rural and the world price of maize as an ad valorem tariff, denoted by t_T^r, so that $p_m^r = p_m^W(1 + t_T^r)$. It is convenient to take 1989 as the base year and calculate any deviations from the base year as a tariff, t^r, that is additional to the tariff ruling in 1989 (denoted by t_0^r). Thus, the wedge between the rural and world maize price is given by $(1 + t_T^r) = (1 + t^r)(1 + t_0^r)$, with the convention that in the base year $t^r = 0$. We linearize the expressions for welfare change around the values observed in 1989, rather than around the free trade values (as is customary) because the free trade values are not known and we would not be able to calculate the linearization constants without a potentially large approximation error. This approach allows us to compute the welfare effects of any tariff, including as special cases a tariff that would produce self-sufficiency (denoted by t_{SS}) and the tariff-equivalent that would produce free trade (denoted by t_{FT}).

We assume tariff revenues are plowed back into the rural sector. (See Newbery 1987 for a discussion of how distributional weights could be incorporated.) This leads to the following aggregate rural budget constraint:

$$(9) \qquad R^r[p_m^W(1 + t_T^r) \ldots] + p_m^W(E_m^r - R_m^r)t_T^r = E^r[p_m^W(1 + t_T^r) \ldots, U^r]$$

where R^r and E^r are the rural revenue and expenditure functions, respectively. Hence, R^r_m is the quantity of maize supplied and E^r_m the compensated demand for maize, so that the second term in parenthesis in the left side of equation 9 is net rural maize exports. Differentiation of equation 9 yields:

$$(10) \qquad E^r_U\, dU = [(R^r_m - E^r_m)p^r_{m,0} + p^W_m(1 + t^r_0)(E^r_m - R^r_m)]\, dt^r$$

$$+ p^W_m(E^u_{mm} - R^r_{mm})\, t^r_T\, p^r_{m,0}\, dt^r + \frac{t^r_0}{(1 + t^r_0)} v_{mr}\, E^r_U\, dU$$

$$= \left[\frac{p^r_{m,0}\,(E^r_{m0}\,\epsilon^D - R^r_{m0}\,\epsilon^S)}{(1 - \frac{t^r_0}{1 + t^r_0} v_{mr})}\right] \frac{t^r_T}{1 + t^r_0}\, dt^r$$

where v_{mr} is the marginal value share of maize in total rural expenditure; ϵ^D and ϵ^S are the price elasticities of demand and supply of maize, respectively; and the subscript 0 refers to base year (1989) values. If we linearize around the base situation, the term in parentheses on the right side of equation 10 becomes a constant, to be evaluated at base year prices and quantities. Integration of equation 10 yields the change in welfare due to a change in the tariff from its value in the base year to any specific target tariff.

$$(11) \quad E^r_U\,[U^r(t^r_T) - U^r(t^r_0)] = \frac{p^r_{m,0}\,(E^r_{m0}\,\epsilon^D - R^r_{m0}\,\epsilon^S)}{(1 - \frac{t^r_0}{1 + t^r_0} v_{mr})} \left[\frac{t^r_0}{1 + t^r_0} t^r + \frac{(t^r)^2}{2}\right]$$

$$= - \zeta^r_0 \left[\frac{t^r_0}{1 + t^r_0} t^r + \frac{(t^r)^2}{2}\right].$$

A formula similar to equation 11 can be derived for urban areas by setting R^r_m (and hence ϵ^S) equal to zero and replacing t^r by t^u:

$$(12) \qquad E^u_U[U^u(t^u_T) - U^u(t^u_0)] = - \zeta^u_0 \left[\frac{t^u_0}{1 + t^u_0} t^u + \frac{(t^u)^2}{2}\right].$$

Expressions 11 and 12 measure the "dead-weight loss" in the rural and urban sectors associated with any tariff and are equivalent to the sum of triangles AWQ and BRZ in panel c and to triangle HGF in panel d of figure 1, respectively. As long as price elasticities are not zero, maize protection has positive welfare costs. These costs increase with the square of the tariff, so increasing protection becomes progressively more costly.

II. Fiscal Costs, Redistributive Impact, and Aggregate Welfare

We now apply the model to an assessment of the fiscal costs and distributive consequences of Mexico's policies on maize pricing and of the welfare gains of moving to free trade.

Table 1. *Estimated Supply and Demand for Maize, 1989*

Category	*Quantity (millions of tons)*
Supply	
Domestic production in rain-fed lands	8.8
Domestic production in irrigated lands	1.5
Imports	2.5
Total	12.8
Demand	
Rural own consumption	3.6
Intermediate use[a]	2.7
Urban consumption	6.5
Total	12.8

a. This includes animal feeds and seeds.
Source: Rivera (1990).

Fiscal Costs

Table 1 gives the 1989 quantities of maize associated with each of the panels in figure 1.[1] The domestic production of 10.3 million tons exceeds rural consumption, so that the rural sector is a net exporter (4 million tons). Because rural exports exceed production on irrigated land, it follows that the subsistence sector as a whole is a net maize exporter. And urban consumption of 6.5 million tons is met by importing 4 million tons from the rural sector and 2.5 million tons from abroad.

The ratio $(p^r_m - p^W_m)/p^W_m$ was 54 percent in 1989; this gives a subsidy of US$72.8 per ton. In 1989 the rural maize price was US$208.5 per ton, and the world maize price (using the yellow maize Gulf price as a proxy) was US$135.5 per ton. Thus for domestic production of 10.3 million tons, the gross produc subsidy is US$749.8 million. But not all of this subsidy is paid by the goverı nent, since part of domestic output is consumed within the rural areas at tl rural price. The net producer subsidy is only US$291.2 million, which is ol ained by multiplying rural exports of 4 million tons by the US$72.8 per tc subsidy; this is area AIJB in panel d of figure 1. The US$458.6 million differen between the US$749.8 million and US$291.2 million gross and net subsidy he tax paid by rural maize consumers. This is made up as follows: a tax (JS$196.5 million paid by intermediate users (2.7 million tons times US$72.8 and a tax of US$262.1 million (3.6 million tons times US$72.8) paid by fin consumers in the rural areas (landless rural workers and subsistence produce who are net buyers).

Large-scale producers sell 1.5 million tons; therefore they appropriate 109. million (or 38 percent) of the US$291.2 million net production subsidy. Th remaining US$182 million (or 62 percent) goes to producers on rain-fed lan who are net sellers. But the ratio $(p^u_m - p^W_m)/p^W_m$ was −37.1 percent in 198

1. The following is based on data presented in Rivera (1990).

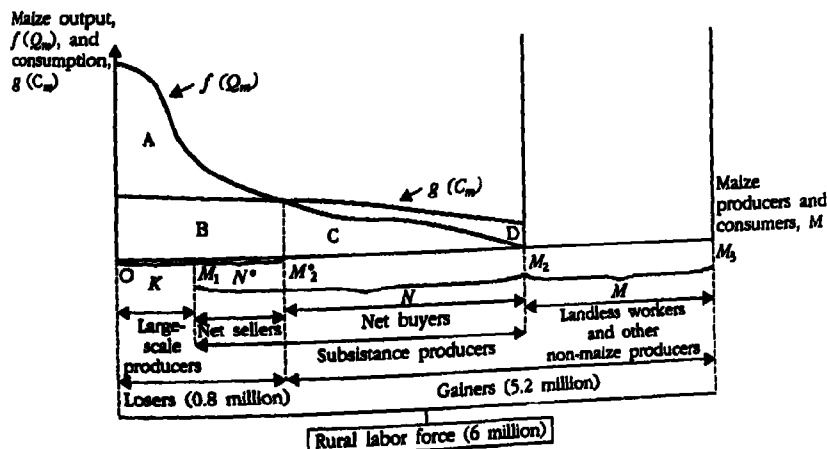
because the average urban price of maize was US$85.2 per ton (the weighted average of the different urban prices; see Rivera 1990 for more details). The subsidy per ton is US$50.37, which when multiplied by the 6.5 million tons of urban consumption gives a subsidy of US$327.40 million (area EFGI in panel d of figure 1).

Redistributive Impact

The redistributive impact works directly through prices and indirectly through the impact of maize prices on labor markets.

Direct price effects. Figure 2 serves as departure point to estimate the direct price effects of lower maize prices on the net income of various groups. The vertical axis measures maize output and consumption by all producers. The horizontal axis lists all maize producers: K large-scale producers and N subsistence producers. Next to N we append a total of M other maize consumers in the rural areas (mainly landless rural workers but also other non-maize producers). Maize producers are listed in order of decreasing output. The underlying assumption is that all large-scale producers on irrigated land have larger maize output than any subsistence producer on rain-fed land.

Let $f(Q_m)$ be the distribution of production, so that the total area under $f(Q_m)$, the sum of A + B + C, is total maize output. The distribution of maize consumption is denoted by $g(C_m)$; area B + C + D is total own-consumption by maize producers. If maize is a normal good and if all producers have the same tastes, $g(\cdot)$ should be a decreasing function, although probably much flatter than $f(\cdot)$. In addition, the area under $g(\cdot)$ may also include some consumption by other rural workers as well as some animal consumption. Masera (1990, table 4.5,

Figure 2. ***Maize Consumption and Production***

p. 126) presents interesting data for this phenomenon from two communities in the state of Michoacan. A similar phenomenon is noted by Andrade (1988, pp. 16–18). All K large-scale producers and N^* subsistence producers are net sellers, while $N - N^*$ subsistence producers are net buyers. Because the subsistence sector as a whole is a net seller, it holds that:

$$(13) \qquad \Sigma_{j=1}^{N^*} (Q^j_m - C^j_m) > \Sigma_{j=N^*+1}^{N} (C^j_m - Q^j_m).$$

Figure 2, however, makes clear that N^* depends critically on the shapes of $f(\cdot)$ and $g(\cdot)$. Hence a condition like inequality 13, by itself, provides insufficient information for policy. In particular, if we take $K + N + M$ to be the total rural population, it follows that there are $K + N^*$ losers and $N - N^* + M$ gainers from the direct price effect of lowering the price of maize. If $f(\cdot)$ is relatively steep, then $K + N^*$ will be a relatively small number, and, while all large-scale producers and subsistence producers as a whole lose, most of the rural population may directly benefit from lower maize prices.

The available data are insufficient to completely trace $f(Q_m)$ and $g(C_m)$ or to determine the exact numbers for K, N, and M. Nevertheless, piecing together estimates in the literature, we find information in two areas. First, the rural population is estimated to be 21.8 million people (or 27 percent of the total); out of this about 6 million are estimated to be the economically active population. In turn, out of the 6 million rural workers, about 2.5 million to 3 million are estimated to be landless, with the remaining 3 million to 3.5 million small producers (Salinas 1990, p. 817; Montanez 1988, p. 684). Montanez and Warman (1985) mention that there are about 2 million subsistence maize producers; the same figure is quoted by Masera (1990, p.39). INEGI (1988, p. 20) also states that there are more than 2 million such producers. However, Masera (1990, table 1.6, p. 40) cites a total of 2.2 million maize producers (of all types). Although some of these figures relate to different years, it seems reasonable to make an estimate of 2.25 million maize producers (out of which 2 million are subsistence and 0.25 million are large scale), about 3 million landless rural workers, and about 0.75 million other rural producers dedicated to non-maize activities. In terms of figure 2 this suggests $K = 0.25$, $N = 2.0$, and $M = 3.75$ million.

Second, data to trace the production and consumption distributions of maize are also scarce. Montanez (1988, p. 679, our translation) states that: "Of all maize producers in the spring-summer cycle, 55 percent do not cultivate more than 2.5 hectares, which yield, on average, 1.35 tons per hectare. Individually, 58 percent obtain, at most, 2.5 tons. It is estimated that 66 percent keep all of its output and that, of the total amount produced, only 50 percent enters the market." Andrade (1988, pp. 11–15) points out that only 5.8 percent of producers obtained more than 10 tons, with average yields of 2.2 tons per hectare; he also shows that 66 percent of all producers harvested between 0 and 2.5 hectares, 32.3 percent between 2.5 and 10 hectares, and only 1.73 percent more than 10 hectares.

These figures imply that out of the 2.25 million maize producers assumed above, only 787,000 (34 percent) are net sellers and that about 135,000 of these (or 6 percent of all producers) account for the bulk of the sales. Thus, the distribution of maize production falls rather steeply at first and then flattens out. With own-consumption requirements falling as well with lower production (less maize required to pay for hired labor and for animal feed), this implies that area D in figure 2 is relatively small. An immediate observation is that the 66 percent of producers who are net buyers probably buy a small share of their total maize consumed, so the benefit they receive from lower maize prices is positive, but probably small. Conversely, there is a group of producers who are net sellers, but whose sales are also relatively small, so their loses from lower maize prices would also not be significant.

This analysis suggests that the indirect wage rate effect of lower maize prices is more important to subsistence producers who are on the margin of being net sellers or net buyers than is the direct price effect. It also indicates that the direct impact of lower maize prices would be strong only for producers in the initial segment of the distribution. A plausible estimate is that this group of significant sellers represents, at most, 15 percent (or 330,000) of all producers (250,000 large-scale producers on irrigated land and 80,000 subsistence producers on rain-fed land). Groups that are definitely net gainers would be those close and to the right of point M_2 in figure 2, or about 3.75 million other rural workers, made up mostly by landless workers. About 1.92 million subsistence producers would lose or gain little.

Indirect labor market effects. To provide an assessment of expression 10, the first step is to calculate the output reduction implied by free trade in maize. A recent survey of econometric estimates of supply elasticities in Mexico shows that the aggregate price elasticity of supply of maize, ϵ^S, is in the order of 1.1 (Nathan Associates 1989, table 4.1). (Since we analyze nonmarginal price changes, we assume that the supply curves of maize are isoelastic.) Unfortunately, these studies failed to distinguish between supply elasticities for irrigated and rain-fed lands, (ϵ^i and ϵ^{rf}, respectively). Since $\epsilon^S = \alpha\epsilon^i + (1 - \alpha)\epsilon^{rf}$, where α is the share of maize cultivated on irrigated land, it is necessary to exogenously assume one of the two individual elasticities (satisfying $\epsilon^i > \epsilon^{rf}$). Lacking additional information, we set ϵ^i equal to 1.5, such that ϵ^{rf} is 1.03 (since $\alpha = 0.15$; see table 1). A move toward free trade in maize would reduce the rural price of maize by, roughly speaking, 50 percent, thus implying a cut in maize output from irrigated land of 0.97 million tons (from 1.5 million to 0.53 million) and of 4.5 million tons from rain-fed land (from 8.8 million to 4.3 million tons).

The second step is to estimate the land and labor released from maize cultivation. Assuming average maize yields from irrigated and rain-fed lands of, respectively, 2 and 1.4 tons per hectare, the output contraction in maize releases 3.21 million hectares of rain-fed land and 0.32 million hectares of irrigated land. Table 2 then implies that 59.1 million worker-days are released from rain-fed

Table 2. *Ratios of Land to Labor, by Crop and Kind of Land*

Kind of land	*Maize*	*Other grains*	*Fruits and vegetables*	*Other*	*Pasture*
Irrigated	51.5	25	165	23	n.a.
Rain-fed	18.4	9	58	18	5

n.a. Not applicable.
Note: Number of worker-days employed per year per hectare of land.
Source: Norton and Solis (1983).

maize cultivation and 16.48 million worker-days from irrigated maize cultivation, so that $|\partial L_m^{rf}/\partial p_m^r + \partial L_m^i/\partial p_m^r| = 75.58$ million worker-days.

The third step is to calculate average labor requirements on non-maize irrigated land and estimate the increase in non-maize employment. Assume the released irrigated land is used to produce other grains, fruits, and vegetables as well as other crops in the same proportions found in table 3. This implies that 62.05 worker-days are required per hectare of non-maize irrigated land. As a result, when the 0.32 million hectares of irrigated land are turned over to other crops, we have that $\partial L_v/\partial p_m^r$ equals 19.85 million worker-days. Consider now what happens to the rain-fed land released from maize. According to agricultural experts, a plausible scenario is that half of the released rain-fed land is devoted to pasture, with the remaining half equally divided between other grains and other crops (and no fruits and vegetables). Table 2 indicates that this use of the released rain-fed land requires 29.6 million worker-days, implying a net release of labor of 26.13 million worker-days. Assuming the average rural worker works 180 days a year in agricultural activities, this translates into 145,000 workers.

Thus, under the assumptions made above and without any other government intervention, free trade in maize would put downward pressure on the rural wage rate. Of course, the wage reduction also depends on migration assumptions, as well as on the elasticities of labor demand in non-maize crops. However, the number of workers released by free trade in maize is small when compared with the total rural labor force (approximately 6 million workers).

Table 3. *Allocation of Land, 1989*
(thousands of harvested hectares)

Crop	*Irrigated*	*Rain-fed*	*Total*
Maize	915	5,553	6,468
Other grains[a]	1,554	1,620	3,174
Fruits and vegetables[b]	1,110	796	1,906
Other[c]	1,367	5,295	6,662
Total	4,946	13,264	18,210

a. Rice, sorghum, wheat, and barley.
b. Including perennial crops and sugar.
c. Cotton, tobbaco, beans, and others.
Source: Direccion General de Estadistica (unpublished data); Secretaria de Agricultura y Recursos Hidraulicos (1984).

Thus, very small migration responses and very small elasticities of labor demand would be required to generate a significant fall in the rural wage rate.

Social Welfare Costs

In this section we look at the aggregate welfare cost of maize policies in 1989 and estimate the welfare gains of free trade in maize. We also evaluate the long-standing Mexican policy objective of maize self-sufficiency. It is difficult to ascertain the economic merit of this objective because its appeal is political and therefore outside the scope of this article. However, we can estimate the economic costs of self-sufficiency.

We apply expressions 11 and 12 to measure the cost of current policies by adding the total welfare gain to both urban and rural groups in moving from the present tariff equivalent, t_0, to the free trade tariff equivalent, $t_{FT} = 1/(1 + t_0) - 1$. Beginning from the base situation, the free trade tariff equivalent would make all producers and consumers face the world price of maize.[2]

To measure the welfare costs of self-sufficiency requires an estimate of the tariff equivalent that would reduce maize imports to zero. This tariff equivalent can be derived from the import demand equation. By definition, at self-sufficiency total maize imports, M, equal zero. To estimate the tariff that yields this outcome, we must also make an assumption about which consumers face the increased maize price. In what follows we assume that self-sufficiency would be reached while maintaining constant the price of maize to urban consumers; this is consistent with the current situation where only rural consumers face the producer price of maize. Thus the self-sufficiency tariff needs to provide enough of a producer subsidy to generate net exports from the rural to the urban areas high enough to offset the effect of urban subsidies. The import demand function for maize is

$$M = E^u_m + E^r_m - R^r_m. \tag{14}$$

The implicit rural self-sufficiency tariff t^r_{SS} needs to solve

$$-E^u_m = E^r_m\,[p^W_m(1 + t^r_0)(1 + t^r_{SS}), \ldots] - R^r_m\,[p^W_m(1 + t^r_0)(1 + t^r_{SS}), \ldots]. \tag{15}$$

Differentiation while ignoring income effects then yields

$$\Delta t^r = -\,(E^r_{m,0}\,\epsilon^D - R^r_{m,0}\,\epsilon^S)^{-1}\,M_0. \tag{16}$$

Table 4 provides the data to carry out the computations. The supply elasticity is much higher than the demand elasticity, something that will affect the relative cost of rural versus urban distortions.

Moving to free trade. Table 5 shows that moving to free trade from the 1989 base year configuration yields welfare benefits of US$154 million per year, most of which is due to a reduction in rural distortions. To put this number in

2. By definition of t, t_{FT} has to satisfy $(1 + t_{FT})(1 + t_0) = 1$.

Table 4. *Basic Maize Statistics, 1989*

	Urban	*Rural*	*Rest of the world*
Basic data			
Consumption, C_m (millions of tons)	6.5	3.6	n.a.
Output, Q_m (millions of tons)	0.0	7.6[a]	n.a.
Net imports M_m (millions of tons)	6.5	−4.0	−2.5[b]
Price (US$ per ton)	85.23	208.50	135.60
Implicit tariff in 1989 (percent)	−37	54	n.a.
Implicit tariff to produce self-sufficiency (percent)	−37	94	n.a.
Basic parameters			
ζ_0 (US$ billion)[c]	0.185	1.99	n.a.
Price elasticity of demand of maize, ϵ^D [d]	0.334	0.334	n.a.
Price elasticity of supply of maize, ϵ^S [d]	n.a.	1.1	n.a.
Marginal income share of maize ν_m (assumed)	0.01	0.01	n.a.

n.a. Not applicable.
a. Output, Q_m, equals gross output *minus* intermediate use (see table 1).
b. World imports of maize minus Mexico's total maize imports.
c. Derived from basic data, above.
d. Nathan Associates (1989).
Source: Authors' calculations.

perspective, assume the 1989 intervention structure is maintained forever. In that case, the costs of the subsidy will rise at the growth rate of gross domestic product (GDP), say 5 percent on average over the medium term. However, future distortionary costs need to be discounted; the relevant discount rate is the marginal real cost of borrowing in foreign markets. Taking recent, post-debt-deal market flotations as a guide, this marginal real cost is estimated at 7.6 percent in real terms, when using a long-term inflation estimate of 5 percent. This makes for a growth-adjusted discount rate of 2.5 percent ([1.076]/1.05 = 1.025). Applying this discount rate yields the results summarized in table 6: the total,

Table 5. *Annual Recurrent Welfare Costs of Maize Price Interventions*
(millions of U.S. dollars)

	Areas of welfare costs		
Intervention	*Rural*	*Urban*	*Total*
Move from free trade to 1989 price structure	122	32	154
Move from 1989 price structure to self-sufficiency	251	n.a.	251
Move from free trade to self-sufficiency	374	32	406

n.a. Not applicable.
Note: Totals may not add because of rounding.
Source: Authors' calculations.

Table 6. *Permanent Growth-Adjusted Welfare Costs of Maize Price Intervention*
(net present value, billions of U.S. dollars)

Intervention	*Areas of welfare costs*		
	Rural	*Urban*	*Total*
Move from free trade to 1989 price structure	4.9	1.3	6.2
Move from 1989 price structure to self-sufficiency	10.0	n.a.	10.0
Move from free trade to self-sufficiency	14.9	1.3	16.2

n.a. Not applicable.
Note: Totals may not add because of rounding.
Source: Authors' calculations.

permanent, but discounted, welfare costs of current maize policies equals about US$6.2 billion, or 3 percent of 1989 Mexican GDP.

Moving to self-sufficiency. Combining data from table 4 with expression 16 yields the self-sufficiency tariff t_{SS} of 0.26. This means that to reach self-sufficiency in maize while maintaining urban consumer subsidies, the rural price needs to be raised by 26 percent above the 1989 level of 54 percent over world prices. Thus the total self-sufficiency tariff would be 94 percent.

The second and third rows of table 5 indicate that the associated welfare costs are substantial. To move from the 1989 configuration to a price configuration that achieves self-sufficiency increases the welfare costs by US$251 million a year. The permanent costs of this policy, using the same growth adjusted discount rate of 2.5 percent, are US$10 billion, or around 5 percent of 1989 GDP. But the real costs are higher, because the relevant comparison should not be the 1989 base year, but free trade; this almost doubles the yearly cost estimate of self-sufficiency to US$406 million. If maintained forever, the total discounted welfare costs in 1989 dollars of self-sufficiency are equal to US$16.2 billion, or 8 percent of 1989 GDP!

Fiscal subsidies and aggregate welfare costs. The analysis shows that in 1989 the government spent US$618.6 million in maize subsidies divided between a net subsidy to the rural areas of US$291.2 million and an urban subsidy of US$327.4 million. Netting out redistributions, we find that the aggregate welfare cost of this policy is US$154 million, divided between US$122 million in the rural areas and US$32 million in the urban areas. For the country as a whole every dollar of subsidy generated only US$0.75 of welfare gain. Thus, the country as a whole loses 25 cents per dollar spent on maize price intervention. For the rural areas every dollar of subsidy generated only US$0.58 of welfare gain, making for a net loss of 42 cents per dollar, while for the urban areas the corresponding amount is 0.90 (a net loss of 10 cents per dollar spent). The difference between rural and urban losses is explained by the fact that in the rural areas subsidies induce a production and a consumption distortion, while in the urban areas the production distortion is absent.

These welfare costs of course depend on the values of the elasticities of supply and demand of maize, parameters that are very difficult to estimate with precision. Nevertheless, it is worth pointing out that the values used are probably on the low side (particularly so with respect to the elasticity of supply of maize),[3] so that these estimates would appear to be a lower bound on the welfare costs of current policies. This observation is reinforced by noting that in the analysis we also have ignored the welfare costs of raising the fiscal revenues required to cover the maize subsidies, a potentially important additional cost (Browning 1987). Unfortunately, there are no studies of the welfare cost of raising fiscal revenues in Mexico. To the extent that this cost is substantial, our estimates of welfare loses would have to be increased accordingly.

III. Why Maize Policies Need to be Changed

A recent study of poverty found that 67 percent of the extremely poor population in Mexico lives in rural areas (Levy 1991). The importance of urban poverty is reduced further if account is made of the fact that not all the extremely poor are equally poor and that the distribution of poverty is not the same across regions. When poverty measures that are sensitive to the depth and distribution of poverty are used, the proportion of people in extreme poverty accounted for by the rural areas in Mexico increases to 76 percent. It is thus very difficult to justify, on poverty alleviation grounds, a subsidy to maize consumption in the urban areas.

Moreover, the reduced urban maize price represents an across-the-board subsidy to all urban consumers. Hence, although poor urban inhabitants benefit, part of the benefit spills over into groups that clearly do not need the subsidy. Since the fiscal cost of this policy is substantial (US$327.4 million in 1989), a targeted program can be equally effective in transferring income to the urban poor while at the same time reducing the fiscal burden and the dead-weight loss. An infra-marginal targeted subsidy (using coupons) also breaks the link between income and the size of benefits received and liberates resources to help the rural poor. Targeted programs are not without problems, however. When benefits are made a function of income, a negative incentive to work is created: if participants realize that benefits fall when their incomes increase they effectively face very high marginal tax rates. From this perspective it is better not to make benefits conditional on income. As poverty programs in Mexico move away from generalized subsidies to means-tested programs, the tension between incentives to work and means-tested targeting will become sharper.

This article has shown that the policy of setting the rural price of maize above the world price is inefficient. The policy precludes workers from being employed in other areas where the value of their marginal product at world prices can

3. The econometric studies surveyed by Nathan Associates (1989) are all based on single equation estimates of maize supply elasticities. We know of no econometric study for Mexico where individual crop elasticities are derived from a profit function approach that incorporates cross-price effects.

ventually be higher, while it induces some of the scarce factor, high-quality ind, to be used in activities for which the value of its marginal product at world rices is lower. The labor allocation is also distorted because, by keeping subsis-ence farmers employed on their own land, higher maize prices reduce the supply f labor and put upward pressure on the rural wage. This is a very indirect and iefficient mechanism for supporting the rural wage rate. The same amount of esources currently spent on maize subsidies could be used for rural infrastruc-ure programs. These programs would provide rural employment and would reate the necessary infrastructure that, over the medium term, is required to pen more earnings possibilities for the rural poor.

We have shown that the distributional effects of the higher maize price are iixed. It directly benefits subsistence producers who are net sellers. But a sub-tantial number of subsistence producers are net buyers; and even while the olicy helps a subset of subsistence producers, larger rents are transferred to irge-scale producers. And finally, landless rural workers are hurt because they iust pay a higher price for the maize they consume.

But perhaps the fundamental problem with the policy of protecting maize roduction is that it focuses on the wrong objective. Rather than helping those /ho produce maize, the objective should be to help poor rural inhabitants, egardless of where they work. Of course reaching the poor directly is noto-iously difficult, especially in rural areas. Indirect methods often need to be pplied, which use an observable variable that one suspects is highly correlated /ith the degree of poverty as a "screening device." We have shown that, because f its differing effects on subsets of the rural poor, the rural price of maize is a ery ineffective screening device. There is in fact a strong presumption that it creens out the relatively poor rather than the relatively rich. With distributional rguments discredited, the central policy objective for maize, therefore, should e to move toward eventual free trade in this commodity.

The FTA provides an ideal opportunity to pursue this objective. It is likely to rovide freer entrance into the U. S. market for other agricultural products in vhich Mexico does have a comparative advantage (such as sugar and fruits and egetables), as well as for the broadest possible range of manufactured products. nsuring secure and sustained access for labor-intensive agricultural and manu-actured products can help ease the transition away from subsistence maize ultivation. If this opportunity is taken, Mexico, unlike many other countries, vill avoid the need to increase protection to agriculture as real incomes increase n the rest of the economy and will be able to provide sustainable increases in iving standards over the medium term to the currently rural poor.

\ppendix. The Relation between the Urban and Rural Labor Markets

To analyze the wage rate effect of reducing the producer price of maize, we irst consider the impact on the rural labor market, assuming there is no migra-ion. We then integrate the urban and the rural labor markets.

The initial rural wage rate, $w^{r,0}$, is obtained by solving

$$\text{(A-1)}\quad L_m^{rf}(w^r, p_m^0; \bar{T}^{rf}) + L_m^i(w^r, p_m^0; T_m^i) + L_v[w^r, p_m^0; (\bar{T}^i - T_m^i)] - L_r - L_p = 0$$

where the superscript 0 denotes the value of the variable before the producer price of maize is reduced, $\bar{T}^i$ is the total endowment of irrigated land, T_m^i is irrigated land allocated to maize, and $(\bar{T}^i - T_m^i = T_v)$ is irrigated land allocated to vegetables, and $\bar{T}^{rf}$ is the total endowment of rain-fed land.

Differentiating the total demand for rural labor, DL_R, and noting that $\partial T_m^i/\partial p_m = -\partial T_v/\partial p_m$, equation A-1 can be rearranged to yield

$$\text{(A-2)}\qquad dDL_R/dp_m = \partial L_m^{rf}/\partial p_m + (\partial L_v/\partial T_v - \partial L_m^i/\partial T_m^i) \cdot \partial T_v/\partial p_m.$$

The first term on the right side of equation A-2 is obviously positive. The second term, however, depends on the comparison between the marginal labor-land ratios in vegetable and maize cultivation on irrigated land. If at the margin vegetables are less labor intensive than maize, the second term will be positive since $\partial T_v/\partial p_m < 0$. Expression A-2 thus tells us that if vegetables are less labor intensive than maize on irrigated land, when p_m falls, the total demand for rural labor unambiguously falls. Conversely, if on irrigated land vegetables are more labor intensive than maize, the change in the demand for labor is ambiguous. Clearly, without migration the contraction in rural labor demand translates directly into lower rural wages. Rural employment stays constant, although its composition changes. Subsistence producers now devote less of their labor to grow maize in their own land, so labor is shed from subsistence maize cultivation. We can calculate the amount of labor that needs to be shed from the rural areas to keep wages constant. Thus, if L_g increased by this amount, the rural wage would remain the same. Differently put, we can calculate the size of a rural employment program that neutralizes the wage effect of reducing the price of maize.

The impact of migration on this outcome depends on the assumptions made about the determinants of rural-urban wage differentials. Before the change in the price of maize, equilibrium wage rates are $w^{r,0}$ and $w^{u,0}$, implying an urban-rural wage differential measured by $\beta^0 = w^{u,0}/w^{r,0} > 1$.

The fall in the price of maize shifts the rural labor demand downward. Without migration the rural wage falls. But this increases the differential between urban and rural wages, which in turn may induce rural migration. If migration occurs until the initial wage differential is reestablished, some rural workers migrate, thus reducing the labor supply in the rural areas and increasing it in the urban areas. Under a constant wage differential assumption part of the wage rate effect of reducing the price of maize is absorbed by lower wages and part by shifting labor from rural to urban areas. Of course, migration need not restore the initial wage differential. But as long as there is some (positive) migration, the fall in the rural wage is mitigated. The counterpart to this is some fall in the urban wage, which in turn increases manufacturing employment. Hence, under

the assumptions stated above and without rural employment programs (ΔL_g = 0), free trade in maize would increase the marginal productivity of capital in manufacturing.[4]

References

The word "processed" describes informally reproduced works that may not be commonly available through libraries.

Andrade Ascencio, J. 1988. "El Autoconsumo de Maiz y Frijol en Mexico." Programa de las Naciones Unidas para el Desarollo. Mexico City. Processed.

Braverman, Avishay, Jeffrey S. Hammer, and Anne Gron. 1987. "Multimarket Analysis of Agricultural Price Policies in an Operational Context: The Case of Cyprus." *The World Bank Economic Review* 1 (2): 337–56.

Braverman, Avishay, Jeffrey S. Hammer, and Choong Yong Ahn. 1987. "Multimarket Analysis of Agricultural Pricing Policies in Korea." In David Newbery and Nicholas Stern, eds., *The Theory of Taxation for Developing Countries*. New York: Oxford University Press.

Browning, Edgar. 1987. "On the Marginal Welfare Cost of Taxation." *American Economic Review* 77 (March): 11–23.

INEGI (Instituto Nacional de Estadistica, Geografia, e Informatica). 1988. *Abasto y Comercializacion de Productos Basicos: Maiz*. Mexico City.

Levy, Santiago. 1991. "Poverty Alleviation in Mexico." PRE Paper 679. The World Bank, Latin America and the Caribbean Country Department II, Washington, D.C. Processed.

Masera Cerutti, O. 1990. *Crisis y Mecanizacion de la Agricultura Campesina*. Mexico City: El Colegio de Mexico.

Montanez Villafana, Carlos. 1988. "Las Condicionantes de la Politica Agropecuaria." *Comercio Exterior* 38 (8): 679–85.

Montanez Villafana, Carlos, and Arturo Warman. 1985. *Los Productores de Maiz en Mexico: Restricciones y Alternativas*. Mexico City: Centro de Ecodesarollo.

Nathan Associates, Inc. 1989. "Comermax: A Multimarket Model of Mexico's Agriculture." Prepared for the Inter-American Development Bank. Washington, D.C. Processed.

Newbery, David. 1987. "Identifying Desirable Directions of Agricultural Price Reform in Korea." In David Newbery and Nicholas Stern, eds., *The Theory of Taxation for Developing Countries*. New York: Oxford University Press.

Newbery, David, and Nicholas Stern, eds. 1987. *The Theory of Taxation for Developing Countries*. New York: Oxford University Press.

Norton, Roger D., and Leopoldo Solís M. 1983. *The Book of CHAC: Programming Studies for Mexican Agriculture*. Baltimore, Md.: Johns Hopkins University Press.

4. Of course, the free trade agreement also shifts the demand for urban labor. If such an agreement results in increased investment (foreign or domestic), demand for urban labor increases, thus increasing urban (and, through migration, rural) wages.

Rivera, Ivan. 1990. "Mexico Agsal II: Regulatory Framework of Food Agroindustries World Bank, Latin America and the Caribbean Country Department II, Washingto D.C. Processed.

Secretaria de Agricultura y Recursos Hidraulicos. 1984. *Anuario Estadistico de la Pro duccion Agricola Nacional 1984*. Mexico City.

Sah, Raaj Kumar, and Joseph Stiglitz. 1987. "The Taxation and Pricing of Agricultur and Industrial Goods in Developing Economies." In David Newbery and Nichol Stern, eds., *The Theory of Taxation for Developing Countries*. New York: Oxfor University Press.

Salinas de Gortari, R. 1990. "El Campo Mexicano Ante el Reto de la Modernizacion *Comercio Exterior* 40 (9): 816–29.

THE WORLD BANK ECONOMIC REVIEW, VOL. 6, NO. 3: 503-528

The Optimal Currency Composition of External Debt: Theory and Applications to Mexico and Brazil

Stijn Claessens

The changes in exchange rates, interest rates, and commodity prices during the past decades have had large impacts on developing countries. Many developing countries have limited access to already incomplete international long-term hedging markets. Thus the question arises whether the currency composition of external debt can be used to minimize exposure to external price risk. Using a utility-maximizing framework, this article shows that, by choosing the optimal currency composition, a country can indeed manage its external exposure. The optimal, risk-minimizing currency composition depends on the relation between export receipts and the costs of borrowings in each currency and on the relations among the costs of borrowings in different currencies. A simple methodology can be used to derive the optimal shares of individual currencies and is applied to Mexico and Brazil. The results show that Mexico and Brazil could have lowered their external exposure to a limited degree by continuously altering the currency composition of their debts. The low correlations between the costs of borrowings and export and import prices make the currency composition of debt a very imperfect hedging tool, and it is likely that hedging instruments directly linked to prices are preferable.

> Other things being equal, a strengthening of the dollar will worsen the terms of trade of net commodity exporters and hence reduce their welfare. For net commodity importers the reverse pattern will hold (Dornbusch 1985, p. 335).
>
> . . . for some developing countries, the fall in the dollar increased the burden of debt relative to their economies (World Bank 1987, p. 49).

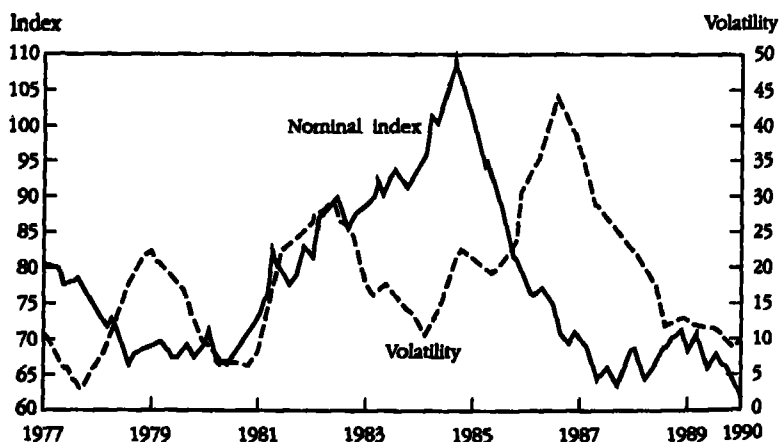
Which of these statements about the effect of exchange rate changes on the welfare of developing countries is correct? Even though placed out of context, both quotations illustrate some of the unresolved issues regarding the effect of exchange rate changes. This article aims to clarify some of these issues and to

Stijn Claessens is with the International Economics Department of the World Bank. He would like to thank Stanley Fischer, Ron Duncan, Kathy Mann, Brian Pinto, Darius Malekpour, the referees, and participants in a seminar at New York University for their comments.

present conceptual and practical guidelines that may help with external det management in general.

The issue of the optimal currency composition of external debt can be ap proached from a narrow or a broad perspective. Given the existence of extern debt, the narrow perspective attempts to determine its optimal compositior Given not only the presence of debt but also the availability of a range of othe international financial contracts, the broader perspective attempts to identify th welfare-maximizing liability structure and the subsequent optimal currenc composition. Here we take the broader perspective.

Developing countries in particular have been affected by large changes in exchange rates, interest rates, and prices of international goods during the pas decades. The degree of uncertainty in these international variables is illustrate in figures 1, 2, and 3. Figure 1 plots the nominal effective U.S. dollar exchang rate and a measure of the volatility of this rate for 1977–90. Exchange rat volatility, which had increased after the movement to floating exchange rates in the early 1970s, did not decline in the 1980s. Figure 2 plots the nominal interes rate most relevant for developing countries from 1965 to 1990 as well as th coefficient of variation (CV) of the interest rate over the preceding 24 months a each point in time. Nominal interest rates have experienced large fluctuations and there have been few periods of tranquillity. Figure 3 plots measures o commodity prices and commodity price volatility during 1962–90. Even thoug part of the price movements can be explained by shifts in the demand fo commodities and supply factors, the large fluctuations in commodity price resulted in large risks to both producers and consumers.

Figure 1. *Nominal Effective U.S. Dollar Exchange Rate: Index and Volatility, 1977-90*

Note: The weights used to create the index of effective exchange rates are the IMF weights. Exchange rate volatility i calculated as the coefficient of variation of the effective exchange rates over the preceding 24-month period. Real effectiv exchange rates show a very similar pattern for level as well as volatility.

Source: IMF, *International Financial Statistics* (various years).

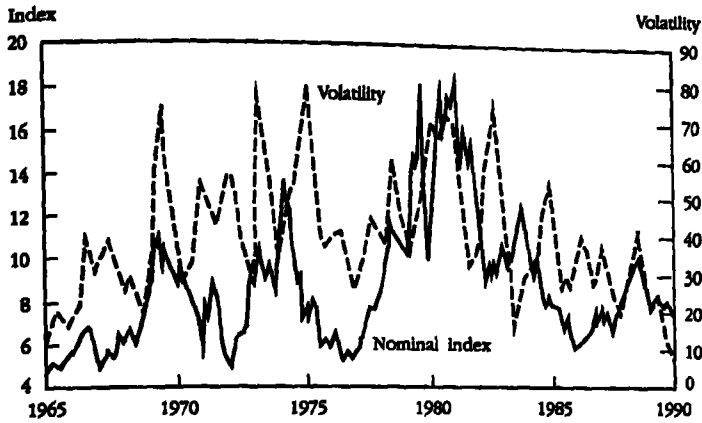
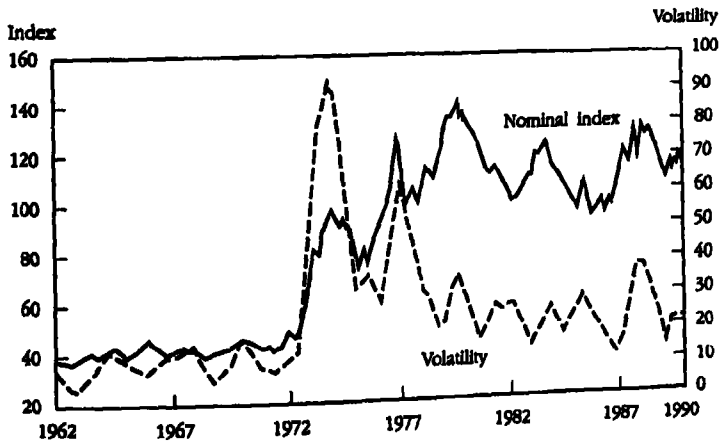
Figure 2. *Nominal Interest Rate: Index and Volatility, 1965-90*

Note: The nominal rate is the London interbank offer rate (LIBOR) on six-month U.S. dollar deposits (period averages in percent per annum) on which much commercial bank debt is indexed. Interest rate volatility is measured as the coefficient of variation over the preceding 24 months. No distinction is made between expected movements and actual deviations from these expectations; overall variability is considered as risk.
Source: IMF, *International Financial Statistics* (various years).

Figure 3. *Nominal Price Index for Developing Countries' Non-Oil Commodity Exports, 1962-90*

Note: The figure plots the commodity price volatility as the coefficient of variation of the IMF index of 34 nonfuel commodity prices for developing countries in the preceding 24 months.
Source: IMF, *International Financial Statistics* (various years).

For industrial countries, external exposures are small, partly because the export and import patterns of these economies are often diversified. In addition, exposures are largely private, and volatilities are thus not an issue for public-budget management. These volatilities have of course been important issues for industrial countries in the larger context of macroeconomic management, but seldom in the narrow context of their financial impact on the public sector. Only a few industrial countries, such as Norway and the Netherlands, have felt a direct financial impact of commodity price and exchange rate movements on their public sector's budget.

But for many developing countries external exposures are large because of a more concentrated export and import pattern. They are also primarily public or quasi-public (as in the case of public external debts and the exports and imports of state enterprises). The high volatilities have had serious implications for the government budgets of these countries and other costs through their disruption of the economy and associated resource misallocations. This is demonstrated by Indonesia, in which the ratio of debt service to exports rose from 8.2 percent in 1981 to 27.8 percent in 1987. More than 65 percent of the increase could be explained by the depreciation of the U.S. dollar after 1985 and the fall in oil and other commodity prices in 1986 and 1987. If commodity prices and cross-currency exchange rates had remained at their end-of-1982 values, the ratio of debt service to exports for all developing countries would have been approximately 17 percent as opposed to the actual ratio of 24 percent in 1987.

The fact that we observe large volatilities in several external prices (exchange rates, interest rates, and commodity prices) and corresponding large changes in measures of debt service burden draws attention to the importance of properly measuring external exposures. Large volatilities of external variables do not necessarily have to affect a firm or a country adversely. The issue is whether and to what extent the interaction between external price movements and overall external transactions affects the firm or country adversely. In order to determine this, a framework for measuring a country's overall (net) exposure to external price uncertainty is needed; the magnitude of external exposures can then be measured.

Once external exposure is quantified, one can try to manage it by using financial instruments to transfer exposures to other parties more able to absorb them. What type of financial instruments are available to manage external risks? Firms in industrial countries have access to and use many financial instruments to reduce and transfer risks. Examples are futures, options, and swaps on currency and interest rates, and more recently on commodities. In many developing countries, however, neither the private sector nor the public sector has used these instruments to the same degree as firms in industrial countries have to shift risks to (international) financial markets in line with comparative advantage. Some hedging instruments—such as currency, interest, and commodity futures and options—are in principle available to most developing countries. However, these have short maturities and, even when rolled over, provide limited hedging

value over longer periods. Furthermore, the problem with rolling-over coverage based on short-term maturing instruments is that this usually implies a large exposure in these instruments. This can imply large margin calls or large option premiums for these instruments, which makes them less attractive for foreign-exchange-constrained countries and reduces them largely to self-insurance instruments. The markets for longer-term hedging instruments—such as currency, commodity, and interest swaps, and especially commodity price-linked instruments—are relatively thin. And these longer-term hedging instruments tend to be unavailable to most developing countries due to institutional, credit, and other constraints.

The presence of large exposures, for which limited or potentially expensive hedging tools are available, raises the question whether a developing country can use alternative means to manage its external exposure. One possibility, of course, is that the country actually diversifies through the sourcing, producing, and exporting of a broader mix of products. However, many developing countries depend heavily on primary exports and may have little room to diversify into other export products as a means to hedge risks. And self-insurance through diversification may not be the most efficient option (because it may not be in line with the country's comparative advantage) and may take a long time to achieve.

A possibly more effective mechanism is to use the currency composition of existing and new external long-term liabilities to reduce external exposure. For example, Kahn (1988) mentions that a number of rescheduling agreements between developing countries and commercial banks have provided currency conversion options for non-U.S. banks. This indicates that, in addition to the possibility of incurring new loans in the currencies of choice, developing countries may be able to alter the currency composition of their existing external debt in rescheduling negotiations, possibly to their own and the commercial banks' advantage. The currency composition of external debt may be able to achieve hedging advantages similar to those of a portfolio of short-term hedging instruments. Changes in external earnings may be offset by simultaneous shifts in the costs of borrowings, but without possible adverse margin calls or premiums. The effectiveness of the currency composition strategy compared with, say, a rolling-over of futures strategy, will depend on the relation among the different external risks. Thus exchange, interest, and commodity price risk management and determination of the currency composition must be integrated.

Several questions follow if a country wants to use the currency composition of external liabilities to manage external exposures. What is the overall objective function to be managed (maximized), and what is the definition of risk that follows from the chosen objective? Which factors play a role in choosing the optimal currency? What kind of rules follow for managing the external liability of a country? The existing literature provides little guidance on these questions. Some objective functions and rules have been suggested for determining the currency denomination of borrowings by a developing country: matching the currency composition of debt with the trade direction or currency composition

of export revenues, matching it with the composition of the basket of currencies with which the domestic currency is pegged or managed, and matching it with the currency composition of the foreign exchange flows into the country (see further Lessard and Williamson 1985). However, most of these rules have not explicitly been related to a specific goal or objective, and the risks have not been explicitly defined. In general, most decision rules proposed have been of an ad hoc nature, and detailed guidelines for implementation have not been developed.

This article develops therefore an integrated model for deriving the optimal currency composition for a small open economy facing external uncertainties. The model uses international portfolio theory and identifies the factors that determine the optimal currency decision. The optimal currency composition for a country to hedge itself against commodity price, interest, and exchange rate movements—and not to speculate on relative exchange rate movements—depends only on the covariances between the effective costs of borrowings and the country's exports and on the covariances among the effective costs of borrowings. A simple operational procedure, which relies on the coefficients of appropriately specified ordinary least squares regressions for calculating the optimal currency shares, is developed and applied to Mexico and Brazil.

I. The Portfolio Model

The optimal currency composition of external debt is determined for a small open economy that acts as a price taker in international goods markets and that faces perfect world capital markets. We will use here a simple, two-period model to determine the optimal portfolio investment and consumption decisions. Alternatively, an intertemporal capital asset pricing model—as in Merton (1971), Breeden (1979), and Stulz (1981)—could be used. In that case the model would have features similar to Krugman (1981), Macedo (1982 and 1983), Fraga (1986), and Stulz (1984). Such a model is derived in Claessens (1988).

It is assumed that the country can be represented by one domestic individual who lives two periods, or, alternatively, the government acts benevolently in choosing the optimal liability composition for its own citizens. This approach is appropriate if the private citizens have limited access to foreign capital markets. For many developing countries this is the case and is confirmed by the small amount of private, nonguaranteed debt compared with total public and publicly guaranteed debt (about 15 percent). In cases where the private sector has more access to foreign capital markets, it may be more appropriate for the government to concentrate on managing its own external liabilities optimally with respect to its fiscal revenues and expenditures.

The country has a fixed first-period endowment and can invest in activities that produce goods for export. In the first period the consumer makes investment, consumption, and borrowing decisions based on its expectations of second-period variables. In the second period, the consumer receives payments on its exports, which it uses to finance debt service payments and imports of

goods. Export receipts and import payments are uncertain because international commodity prices are uncertain. The consumer maximizes its welfare function defined over this and next period's consumption of each good, c_i, $i = 1, \ldots K$.

The country does not face any borrowing or lending constraints and can denominate its liabilities (borrow) and invest its wealth (lend) in N currencies. The amount borrowed or lent in currency j is denoted by B_j, which is a representative element of **B**, the $N \times 1$ vector of demand for foreign bonds. The effective cost (or returns) of B_j in terms of the domestic currency depends on the foreign interest rate and the exchange rate. Each of the N foreign interest rates, r_j^*, is assumed to be uncertain. Similarly, each of the N exchange rates, e_j, defined in terms of units of domestic currency per unit of foreign currency (for example, pesos per dollar), is also uncertain. The effective (gross) interest rate on B_j is defined as R_j^*, and $\mathbf{R}^*$ is the $N \times 1$ vector of the expected effective costs on the N foreign loans. R_j^* is thus given by the (gross) foreign interest rate r_j^* times (one plus) the rate of depreciation of the domestic currency in relation to currency j, that is, $(1 + r_{j,1}^*)(e_{j,2}/e_{j,1})$. It is assumed that the effective costs of liabilities denominated in the currency of each foreign country do not necessarily equal each other. And the standard finance assumption of no transaction costs is made.

Due to certain barriers, domestic investors are prevented from investing in foreign firms and stocks. This restriction prevents domestic investors from using foreign equities as hedging instruments against unanticipated changes in external prices. Foreign investors cannot own domestic firms. We assume that there are no nontraded domestic assets (such as human capital). (See Svensson [1989] for the case of nontraded assets.)

Domestic assets that are traded are domestic bonds, denoted by D, which are in zero net supply, with a gross interest rate of R, and shares of export firms. Export goods are produced by domestic, profit-maximizing firms whose shares are not traded abroad. In the first period investment levels, denoted by I_h, $h = 1, \ldots L$ are determined for production of each of the L export goods in the second period. The production functions, denoted by $Q(I_h)$, exhibit decreasing returns to scale, and $\mathbf{Q}(\mathbf{I})$ denotes the vector of second-period outputs. First-period prices of all goods are normalized at one, and second-period prices in the domestic currency for the export goods are denoted by P_h^*. In the second period the country will receive payments per unit of exports in domestic currency equal to P_h^*. The word "price" should be interpreted broadly here and includes all factors that determine the value of one unit of an export good. The exact currency in which the payments are received is immaterial to our results: what matters is that the payments have a stochastic unit value in terms of domestic currency. Second-period import good prices are P_i.

The first- and second-period asset flow constraints for the representative consumer are defined in equations 1 and 2, respectively:

$$\sum_i^K c_{1,i} = \sum_j^N B_j - \sum_h^L I_h + D \tag{1}$$

$$(2) \qquad \sum_i^K p_i c_{2,i} = -\sum_j^N R_j^* B_j + \sum_h^L P_h^* Q(I_h) - RD$$

where first-period exchange rates are normalized to one.

It is assumed that the (representative) domestic investor maximizes an expected utility function, which is von Neuman-Morgenstern and depends only on the consumption of the K commodities. The expected utility function of the investor is given by

$$(3) \qquad U\left(\prod_{i=1}^{k} c_{1,i}^{\alpha_i}\right) + \beta E\left[U\left(\prod_{i=1}^{k} c_{2,i}^{\alpha_i}\right)\right]$$

where Π is the standard multiplication function, E is the expectation operator conditional on all information available at time 1, and β is the factor of time preference. The utility function is characterized by constant expenditure shares: the investor will always spend a share α_i of total expenditures on good i. This allows us to represent the objective function in terms of one composite good for which we will use the notation C_i, $i = 1,2$, with average price P (geometrically weighted with weights α_i). The investor will maximize utility subject to the constraints on income.

$$(4) \qquad \max_{B_j, D, I_h} U(C_1) + \beta E[U(C_2)]$$

$$\text{subject to } C_1 - \sum_j^N B_j + \sum_h^L I_h - D = 0$$

$$\text{and } PC_2 + \sum_j^N R_j^* B_j - \sum_h^L P_h^* Q(I_h) + RD = 0.$$

To simplify the solution for the optimal amounts borrowed in each foreign currency, invested in domestic bonds, and held in first-period investment levels, we assume that the utility function U is quadratic, that is, $U = aC - (b/2)C^2$. For the more general case, see Claessens (1988) or Svensson (1988). We define $\mathbf{V}_{rr}$ as the $N \times N$ variance-covariance matrix of the effective costs of foreign borrowings deflated by the consumer price index and $\mathbf{V}_{rp^*}$ as the variance-covariance matrix of those costs with the changes in the unit values of the export goods (again expressed in domestic currency and deflated by the consumer price index). After imposing that the domestic market for bonds clears, the optimal borrowings are, where $\mathbf{1}$ is a $N \times 1$ vector of ones (see further the appendix):

$$(5) \qquad \mathbf{B} = \mathbf{V}_{rr}^{-1} \left\{ \frac{\mathbf{E}[(1/P)\mathbf{R}^*] - \mathbf{1}\dfrac{U_1'}{\beta E(U_2')}}{-\dfrac{b}{E(U_2')}} \right\} + \mathbf{V}_{rr}^{-1}\mathbf{V}_{rp^*}\,\mathbf{Q}(\mathbf{I}).$$

Equation 5 implies that the demand for foreign bonds consists of two mutual funds. The first fund is the speculative portfolio:

$$B^s = V_{rr}^{-1}E[(1/P)R^* - 1\tau]$$

$\tau = U_1'/(\beta E(U_2'))$ is the effective rate of intertemporal substitution. The
›sition of the speculative portfolio of foreign bonds depends on the ex-
l effective real costs of excess borrowings. In equilibrium these costs will
d on the real rate of return on investment, adjusted for risks, and the
e of the variance-covariance matrix of effective real costs. The demand for
nd depends on a risk aversion parameter, γ, where $\gamma = -b/E(U_2')$. The
the aversion against risks, the higher the γ, and the lower the amount
ved on account of the speculative fund.

second portfolio is the minimum variance hedge:

$$B^m = V_{rr}^{-1}V_{rp^*}Q(I)$$

independent of the level of risk aversion. The investor will borrow in
n currencies to insulate against changes in the domestic currency value of
receipts from exports. These changes are caused not only by movements
domestic exchange rate at which export receipts are converted into domes-
rency, but also by movements in the unit values of export goods. Since
currency movements can be related to unit value movements, foreign bor-
gs can serve as a hedge, and the stronger the correlation between domestic
icy values and the cost of borrowings the larger the demand for foreign
wings.

equation for the optimal amount of loans is specified with respect to the
ve, real cost of foreign liabilities, relative to the domestic borrowing costs.
atively, the portfolio can be written in nominal terms and then split into
funds: a nominal speculative fund, a minimum variance fund, and a price
(see the appendix). In addition, we need to assume that the distribution of
ces is lognormal. The nominal fund will be $V_{rr}^{-1}R$, and the minimum
ce fund will be $V_{rr}^{-1}V_{rp^*}Q(I)$, and the matrices are now based on the
ial costs of borrowings and the nominal unit values of exports. The price
fund will be $-V_{rr}^{-1}V_{rp}$, where V_{rp} is the vector of covariances between the
ve costs of borrowings and the consumer (import) price index. The de-
for the price level hedge will be $-(1 - 1/\gamma)$ (see the appendix).

useful to look at the price level hedge when we assume that the foreign
t rates are constant and thus that the uncertainty in the effective foreign
t rates is only a result of exchange rate uncertainty. This implies that the
ce-covariance matrix of effective costs (V_{rr}) would be equal to the
ce-covariance matrix of exchange rates, and the vector of covariances
n the effective costs of borrowings and the consumer price index (V_{rp}) is
o the vector of covariances between exchange rates and consumer prices.
ve look at the price level hedge in two cases. In the first case changes in the
of imported goods are perfectly correlated with the exchange rate, that is,
v of one price holds for imported goods. Holding foreign bonds provides
perfect hedge against unanticipated changes in the domestic price of
n goods. The demand for each foreign bond will be determined by the

share of total consumption expenditures spent on the imports from each currency area. In the second case purchasing power parity holds. Foreign bonds are perfect hedges against unanticipated domestic price movements as exchange rate movements offset perfectly relative price movements. With no unanticipated inflation abroad there will be no demand for foreign bonds on account of the price level hedge.

To summarize, the model outlined above indicates that the optimal external liability composition of a country depends on the following factors:

- The expected costs of borrowing in each of the foreign currencies relative to the domestic cost of funds
- The variance-covariance matrix of the expected costs of borrowing in each foreign currency
- The variance-covariance matrix of domestic goods prices and expected costs of borrowings in each foreign currency
- The shares of consumption expenditures spent on the different goods in the country
- The export receipts of the country
- The vector of covariances between export receipts and expected costs of borrowings in each foreign currency
- The level of risk tolerance in the country.

When purchasing power parity holds and foreign prices are (relatively) stable, the rules for the optimal currency composition of external liabilities can be further simplified. A country may be very risk averse and want to hedge itself against commodity prices and interest and exchange rate movements. Or it may take the view that the expected real costs of borrowing in each of the foreign currencies are equal to the domestic real cost of funds (that is, it may not want to speculate on the relative costs of different instruments). In these cases the optimal composition will depend only on the vector of covariances between export receipts and the expected costs of borrowing in each foreign currency, and the variance-covariance matrix of the expected costs of borrowing in each foreign currency.

II. Practical Rules for the Currency Composition of External Debt

From the theoretical analysis it follows that if a country does not want to take an active view on exchange and interest rate movements—and associated costs of borrowings—or if it is relatively risk averse, then its optimal *net* borrowing portfolio (gross borrowings minus foreign assets, that is, reserves) will be the risk-minimizing hedge portfolio. This is a very familiar result from simple one-period mean-variance hedging models such as those used to determine the optimal amounts of futures to buy or sell to hedge an exposure (see, for example, Gemmill 1985). Techniques introduced by Adler and Dumas (1980) and later on

refined by others for the operational measurement of the economic exposure of a firm to external risks are also similar. We will pursue the latter similarity further.

Economic exposure to external risks for a firm has been defined in terms of the sensitivity of its objective function with respect to unanticipated changes in external variables. The firm's objective function may be defined in terms of the net present value of future expected cash flows or in terms of near-term cash flows or profits. When the hedging instruments available are foreign borrowings, the operational way of measuring exposure as a cash flow sensitivity can then be reduced to measuring the covariance between cash flows and the effective costs of borrowings relative to the variance of total cash flows. In other words the measure is the ordinary least squares (OLS) regression coefficient of cash flows on the relevant effective costs of borrowings: $\mathrm{cov}(CF_t, R_t)/\mathrm{Var}(CF_t)$, where CF_t is the cash flow in terms of domestic currency in period t and R_t the relevant effective cost of borrowings at time t. The absolute values of these estimated coefficients provide then the minimum variance hedging quantities. These are the amounts of foreign borrowings (multiple exposure measures follow if one regresses the cash flows on multiple cost of borrowings). A related exposure measure is discussed by Oxelheim and Wihlborg (1987).

The portfolio model above is consistent with this approach because it indicates that—when the expected foreign and domestic real costs of funds are equal and purchasing power parity holds between foreign and domestic goods—the optimal composition is simply determined by the minimum variance fund $(V_{\pi\pi}^{-1} V_{\pi p} \cdot Q(I))$. In the practical approach we consider neither the relative costs of funds in different currencies nor the benefits of using foreign borrowings to hedge against domestic price uncertainty. It is unlikely that a developing country will be able to successfully exploit (speculate) at an acceptable risk the small differences between costs of funds. The international financial capital markets at large are better equipped for this and will assure that these differences remain small.

Domestic price uncertainty is more often the result of endogenous government policy choices than of exogenous influences; consequently it is unlikely that it can be hedged using foreign borrowings. It is not assumed that the law of one price holds among foreign goods (that is, P_t is not necessarily equal to $e_h P^*_{i,h}$ for all i and h, where $P^*_{i,h}$ is the price of traded good i in terms of foreign currency h) and foreign exchange rates can influence the prices of foreign goods. We leave the precise mechanisms through which nominal foreign exchange rate movements influence prices of foreign goods unspecified. For the effects of currency movements on the behavior of absolute and relative prices, see Giovannini (1986), Dornbusch (1987), and Varangis and Duncan (1987).

More generally, the risk-minimizing borrowing portfolio would be based on the relation between the ability of a country to service its debt and the effective cost of debt service in each of the relevant borrowing currencies. The main problem with this approach is measuring the ability of a country to pay, some-

thing that has been the subject of extensive research in the sovereign debt literature. In addition, ability and willingness to pay may differ, thus resulting in actual payments that may be determined by a bargaining game between a debtor and its creditors (see Eaton [1990] for a survey). In that case the factors determining the payment behavior belong on the right side of the OLS equation.

We suggest here that total exports of a country is the relevant measure to hedge. (Healy [1981] uses a similar simple regression technique for the optimal diversification of foreign exchange reserves.) Total exports measured in domestic currency are thus regressed on the effective costs of foreign funds (both expressed as percentage changes), and the coefficients are used to calculate the optimal liability portfolio. The equation to be estimated becomes then

$$(8) \qquad \ln(P^*Q)_{t+1} - \ln(P^*Q)_t = \sum_i \beta_i \left\{ \ln\left[(1 + r_{i,t})\left(\frac{e_{i,t+1}}{e_{i,t}}\right)\right] - \ln\left[(1 + r_{i,t-1})\left(\frac{e_{i,t}}{e_{i,t-1}}\right)\right]\right\} + u_t.$$

Further disaggregation, for example by commodity groups or direction of trade, can also be used to calculate the optimal portfolio shares if one expects that the disaggregated relationships are more stable. For instance, prices of some commodities tend to have a close relationship with a single currency because the supply (or demand) tends to come from (or go to) that currency area. For example, the price of coniferous timber products has been closely associated with the Nordic countries' currencies because they are large suppliers and influence the dollar price of timber. Similar relationships may exist for other commodities (see further Lessard and Williamson 1985).

III. Applications to Mexico and Brazil

In applying the model, the objective is to find the currency composition of net liabilities (gross liabilities minus reserves) that minimizes the variability in domestic currency of export earnings net of foreign liability debt service. We assume that at the beginning of each period new net liabilities are incurred, which are then paid off at the end of the period. We have thus a rolling portfolio of liabilities. If the composition of the optimal portfolios obtained is now stable over time, then little rebalancing of the portfolio would be required, and the rolling portfolio would essentially mimic a portfolio of long-term liabilities that has debt service payments falling due each period. We would then be justified to use the currency composition of long-term liabilities (which is inherently difficult to change over short periods) to hedge the short-term exposure arising from changes in export earnings. Large reductions in overall variability would indicate that the currency composition of external debt can be an effective tool for risk management, especially since it would avoid the drawbacks of rolling over short-dated instruments (margin calls and premiums).

We use monthly data for 1973–89 for Mexico and for 1973–86 for Brazil. In accordance with the model, we use as the dependent variable Mexico's and Brazil's total exports, expressed in domestic currency. (All data are from the International Monetary Fund's *International Financial Statistics*.) As possible borrowing currencies we choose the U.S. dollar, Japanese yen, and German deutsche mark (DM). These three currencies were selected for the borrowing portfolio because Mexico and Brazil's current debt is largely denominated in these currencies (respectively, 70 and 72 percent U.S. dollar, 15 and 11 percent DM, and 8 and 10 percent yen); the DM and other European currencies are highly correlated, and the three chosen currencies are therefore good proxies for the diversification available in international currency markets; and these countries' future access to financial markets is likely concentrated in these currencies (or currency blocks).

The estimations were based on the (monthly) percentage change in the costs of borrowing in the different currencies expressed in local currency. The foreign interest rates used were the monthly quotes of the respective six-month LIBOR rates, which were then multiplied with the exchange rate depreciation over the next month. Table 1 presents some data on the average annualized costs in pesos during 1973–89 and in cruzados during 1973–86. The variability in the local cost of foreign borrowings was quite high. The variability of foreign costs can be compared with the monthly coefficient of variation of export earnings in local currency: 62 percent for Mexico and 233 percent for Brazil.

The estimate for the optimal portfolio for the whole sample period was based on regression equation 6. The estimates for the optimal amount borrowed in each currency lead to the portfolio compositions for Mexico and Brazil given in table 2. The absolute amount to be borrowed as a percentage of export earnings is 0.25 percent for Mexico and 5.62 percent for Brazil. The small fractions reflect the fact that the variability in export earnings is much less than the variability in effective costs of borrowings and that the correlation between the two is low. The results indicate that, of the absolute amount to be borrowed,

Table 1. *Average Annual Costs of Borrowings*
(percent)

	Cost in domestic currency			
Currency of debt	*Mean*	*Coefficient of variation*	*Minimum*	*Maximum*
Mexico, 1973–89 (pesos)				
U.S. dollar	46	256	−156	1,133
DM	46	266	−138	1,059
Yen	48	249	−144	1,100
Brazil, 1973–86 (cruzados)				
U.S. dollar	70	90	6	475
DM	70	106	−92	467
Yen	72	104	−105	444

Source: Author's calculations.

Table 2. *Optimal Portfolio Compositions for Mexico and Brazil*

	Percentage of total amount borrowed in			*Amount borrowed as a percentage of export earnings*	R^2	*Durbin-Watson statistic*
	U.S. dollars	*DM*	*Yen*			
Mexico, 1973–89	111.0	16.3	−27.4	0.25	0.068	2.00
Brazil, 1973–86	98.0	3.4	−1.5	5.62	0.053	2.04

Source: Author's calculations.

Mexico should borrow more than 100 percent in U.S. dollars, borrow some additional DM, and hold some of the foreign funds in yen reserves. Brazil should also borrow U.S. dollars and DM and hold yen in reserves.

If Mexico and Brazil had borrowed and invested their reserves in this fashion during 1973–89, the variance of monthly exports net of debt service would have been slightly lower compared with no borrowings and no hedging. The small reduction in uncertainty reflects the fact that the fits of the estimated equations are relatively poor (R^2 of only 0.068 and 0.053 for Mexico and Brazil, respectively) and that the underlying relationships between export earnings and costs of borrowings are not stable over time. The latter implies that portfolios should be more frequently adjusted and thus that the hedging effectiveness of a constant portfolio is limited.

The results so far were calculated using historical data and then applied over the previous sample period. The results therefore amount to an in-sample test. However, such historical comparisons cannot indicate the effectiveness of the strategy as a planning tool. After all, in reality, the country would only have had information up to time t with which to determine the composition for time $t + 1$. To show whether or not this technique is effective in terms of risk minimization, we perform a number of "out-of-sample" tests. We construct portfolios for period $t + 1$ that are based on a sample of observations up to time t. The sample is then updated to include time $t + 1$ information, portfolios are constructed for period $t + 2$, and so forth. The relative effectiveness of this strategy of rolling hedges is then calculated as the risk reduction realized over the whole hedging period compared with the risk reduction realized by the other strategies.

Specifically, we calculated optimal portfolios for Mexico and Brazil for each quarter using the last 48 monthly observations up to that quarter. This results in 51 portfolios for Mexico and 39 for Brazil. Table 3 summarizes the results. Annex tables A-1 and A-2 list the individual portfolios, indicating the mean square residual, the (adjusted) R^2, and the Durbin-Watson statistic. For some individual portfolios the total borrowing amount as a percentage of export earnings is negative, thus implying that a net lending strategy would have been preferred at some points in time.

The summary table indicates that the standard deviation of the portfolio shares is quite large. Portfolio shares vary between −3,622 and 2,997 percent for Mexico and between −123 and 202 percent for Brazil. Some of the large

individual portfolio shares are caused by the very low absolute amount to be borrowed in a specific quarter, which inflates the individual portfolio shares. For example, the total borrowing amount as a percentage of export earnings of portfolio 37 for Mexico is only 0.000087.

Some of the outliers in the case of Mexico are also caused by the two large discreet valuations that occurred in 1982, which distort the estimation for all portfolios that include these data points. This is for instance reflected in the amounts to be borrowed for portfolios 23 through 39 (1982–87 through 1986–87, which thus include the 1982 data), which are all negative. To control for this, we have also excluded from the sample the negative sums. We also solved for the optimal shares while imposing some constraint on the absolute amounts to be borrowed, for example, a certain ratio of debt service to exports. The results for both approaches were, however, not significantly better. Similar explanations exist for Brazil for portfolios after 1983-10. Excluding these periods, the estimates for the portfolios indicate a consistent strategy of borrowing in U.S. dollars combined with some minor DM or yen borrowings or reserve holdings. The absolute amounts to be borrowed are lower than the actual borrowings by these two countries and are closer to the net transfers these countries pay on their debts, expressed as a fraction of exports, consistent with the idea that the portfolios are rolled over.

With these individual portfolios, we performed the following out-of-sample test. For any portfolio one can calculate the residual in each of the next three months that was not hedged as $u_t = \Delta E_t - \sum_i \beta_i \Delta R_{i,t}$, where ΔE_t is the realized monthly percentage change in export earnings, $\Delta R_{i,t}$ is the realized monthly

Table 3. *Summary Statistics of Optimal Portfolios for Mexico and Brazil*
(percent)

Variable	*Mean*	*Standard deviation*	*Value*	
			Minimum	*Maximum*
Mexico				
Share of debt in				
U.S. dollars	257	499	−89	2,997
DM	63	174	−72	940
Yen	−309	679	−3,622	29
Total borrowing amount as a percentage of export earnings	2	3	−1	8
Brazil				
Share of debt in				
U.S. dollars	56	93	−97	202
DM	−15	23	−123	7
Yen	2	10	−8	38
Total borrowing amount as a percentage of export earnings	2	11	−20	17

Note: For Mexico there were 51 observations; for Brazil there were 39.
Source: Author's calculations.

percentage change in the costs of borrowing (for i = U.S. dollar, DM, and yen, all expressed in local currency), and β_i are the shares, which change every three months. A new portfolio is then used for the next three months to calculate the residual, and so forth. This allows us to compare the effectiveness of a particular borrowing strategy with any other borrowing strategy.

This was done here by comparing the strategy where portfolio shares are adjusted every quarter according to the estimates derived from the observations during the previous 48 months—here called the rolling strategy—with three alternative strategies. The first alternative strategy was Mexico's and Brazil's actual debt composition in 1988 (reported above); the second one was the composition that was historically optimal for 1977–89 and 1977–86 (89.8 and 126.6 percent U.S. dollar, 11.5 and −25.0 percent DM, and −1.4 and −1.6 percent yen, respectively, for Mexico and Brazil); and the third one was the composition that was historically optimal for 1973–89 and 1973–86 (reported in table 2). For all three alternatives only the composition of debt was changed. The total amount to be borrowed (or loaned) was assumed to be equal to the rolling strategy (otherwise using the optimal amounts of the historically optimal portfolios would amount to an in-sample test).

For both countries the residuals of all four portfolios were calculated; in total there were 152 residuals for Mexico and 117 for Brazil. The results are shown in table 4. For Mexico the results are very encouraging. The rolling strategy achieved approximately a 46 percent lower coefficient of variation of residuals compared with the actual borrowing strategy. The compositions that were historically optimal for 1977–89 and 1973–89 achieved less or no risk reduction compared with the actual portfolio, and the rolling portfolios outperformed all three portfolios. For Brazil, the results are less encouraging. The rolling strategy achieved only a 2 percent lower coefficient of variation than the actual borrowing strategy. Much risk reduction was achieved with the rolling portfolio compared with the 1977–86 historically optimal strategy (a 69 percent reduction), but not compared with the 1973–86 historically optimal strategy, which per-

Table 4. *Risk Reduction Achieved with Optimal Portfolios*
(compared with the actual portfolio)

	Mexico		*Brazil*	
Portfolio	*Coefficient of variation*	*Reduction (percent)*	*Coefficient of variation*	*Reduction (percent)*
Actual	686.0	n.a.	367.0	n.a.
Rolling strategy	371.9	46	358.9	2
Optimal composition for				
1977–89	565.8	18	n.a.	n.a.
1973–89	916.2	−33	n.a.	n.a.
1977–86	n.a.	n.a.	1,153.0	−214
1973–86	n.a.	n.a.	356.0	3

n.a. Not applicable.
Source: Author's calculations.

formed overall the best. For both countries, all strategies did much better than a strategy using the actual amounts as well as the actual shares in which the countries borrowed (statistical results are not reported here).

None of the strategies performed better than a strategy where both composition and amounts were determined in an historically optimal way. This may be expected since historical portfolios will always lead to risk reduction given the benefits of hindsight. The instability in the covariances between changes in costs of borrowings and exports implies that planned portfolios will always result in less hedging. Only with more stability can better results be expected from planned portfolios compared with historical ones. This is, however, not a fair comparison, because, as noted above, in-sample tests of effectiveness of hedging are misleading. Other out-of-sample tests were also performed. One of these was based on a rolling hedge strategy using the past 24 months of observations. This led to much larger standard deviations in borrowing shares, a result of the fact that the sample period for the estimation was halved. Compared with the actual and historical optimal composition this strategy led to little or no risk reduction.

IV. Conclusions

This article shows that an integrated analysis of the external risks an economy is facing is necessary to measure external exposures correctly. The article uses international portfolio theory for such an analysis and finds that in particular exchange and commodity price exposure need to be integrated. A developing country wishing to implement more active management with respect to the currency composition of its net liabilities will need to consider its objective function very carefully. Nominal dimensions (such as direction of trade flows and currency composition of cash-flows receipts) do not necessarily provide correct indications for the optimal currency composition of debt. Determining the optimal currency composition requires a careful empirical investigation of relationships between cross-currency exchange rates and indicators of the country's external account, such as terms of trade, exports, and the non-interest current account. Determining the optimal liability structure requires facts such as the historically observed inverse relationship between the value of the dollar and dollar commodity prices. This relationship indicates for instance that non-dollar currencies can perform a hedging function for (primary) commodity exporters because commodity prices, and thus export earnings, increase when nondollar currencies rise in value and vice-versa.

The article applies a reduced form of the analytical model developed to Mexico and Brazil and finds some, albeit limited, reduction in external exposures from constructing optimal, rolling portfolios. The results indicate, however, that portfolios are unstable over time as a result of time-varying covariances. Thus transactions costs can significantly reduce the benefits of adjusting portfolios frequently. Results obtained elsewhere (Kroner and Claessens 1991) indicate

that even when using Generalized AutoRegressive Conditional Heteroskedasticity (GARCH) to deal with time-varying covariances—and to achieve less time-varying portfolios—instability remains an issue and that using the currency composition of long-term debt in a cost-effective way to hedge against real shocks will be difficult.

In general, the results indicate that the currency composition of debt is an imperfect hedging tool against external price uncertainty. Other types of contingent contracts (such as commodity price linked instruments) are likely better able to transfer risk from developing countries to the international capital markets. Unfortunately, the longer-term spectrum of commodity risk management instruments is not yet well developed for many commodities of interest to developing countries and will in any case be difficult to access because of credit constraints. In principle, developing countries should then also be using the short-dated commodity-linked markets, such as options, forwards, and futures, to hedge directly against price risks. These will provide the countries with some risk reduction benefits when longer-term markets become available or the countries achieve some real diversification. Unless strong relationships are found between the prices of export and import products, the currency choice of external debt may for many developing countries be largely immaterial as it relates to reducing overall exposure to external factors.

Appendix. Derivation of the Optimal Portfolio Model

For simplicity we will derive first the optimal portfolio amounts in case only one foreign bond and one domestic investment opportunity are available. The generalization to N foreign bonds and L investment opportunities will then follow easily.

The Simplified Case of One Foreign Bond and One Domestic Investment Opportunity

From the maximization problem in equation 4 the following first-order conditions with respect to the amount of investment, foreign bonds, and domestic bonds can be derived.

$$(\text{A-1}) \qquad 0 = -U_1' + \beta E\left(U_2' \frac{P^* Q'}{P}\right) \qquad (\text{w.r.t. I})$$

$$(\text{A-2}) \qquad 0 = -U_1' + \beta E\left(U_2' \frac{R^*}{P}\right) \qquad (\text{w.r.t. B})$$

$$(\text{A-3}) \qquad 0 = -U_1' + \beta E\left(U_2' \frac{R}{P}\right) \qquad (\text{w.r.t. D})$$

Given the assumption of a quadratic utility function, we can rewrite the first-order conditions as

$$\text{(A-1}'\text{)} \quad 0 = -U_1' + \beta\Big[E(U_2')E\Big(\frac{P^*Q'}{P}\Big) - b\mathrm{cov}\Big(\frac{P^*Q'}{P},\frac{P^*Q}{P}\Big) + b\mathrm{cov}\Big(\frac{P^*Q'}{P},\frac{R^*}{P}\Big)B + b\mathrm{cov}\Big(\frac{P^*Q'}{P},\ \frac{R}{P}\Big)D\Big]$$

$$\text{(A-2}'\text{)} \quad 0 = -U_1' + \beta\Big[E(U_2')E\Big(\frac{R^*}{P}\Big) - b\mathrm{cov}\Big(\frac{R^*}{P},\frac{P^*Q}{P}\Big) + b\mathrm{cov}\Big(\frac{R^*}{P},\ \frac{R^*}{P}\Big)B + b\mathrm{cov}\Big(\frac{R^*}{P},\frac{R}{P}\Big)D\Big]$$

$$\text{(A-3}'\text{)} \quad 0 = -U_1' + \beta\Big[E(U_2')E\Big(\frac{R}{P}\Big) - b\mathrm{cov}\Big(\frac{R}{P},\frac{P^*Q}{P}\Big) + b\mathrm{cov}\Big(\frac{R}{P},\ \frac{R^*}{P}\Big)B + b\mathrm{cov}\Big(\frac{R}{P},\frac{R}{P}\Big)D\Big].$$

Imposing that in equilibrium the amount of domestic bonds will be in zero net supply, equation A-3′ can be solved for the domestic equilibrium nominal interest rate R as

$$\text{(A-4)} \quad E\Big(\frac{R}{P}\Big) = \frac{U_1'}{\beta E(U_2')} + \frac{b}{E(U_2')}\mathrm{cov}\Big(\frac{R}{P},\frac{P^*Q}{P}\Big) - \frac{b}{E(U_2')}\mathrm{cov}\Big(\frac{R}{P},\frac{R^*}{P}\Big)B.$$

After some manipulation equation A-2′ can be used to find the expression for the optimal amount of foreign bonds as

$$\text{(A-5)} \quad B = \frac{E\Big(\frac{R^*}{P}\Big) - \frac{U_1'}{\beta E(U_2')}}{-\frac{b}{E(U_2')}\mathrm{VAR}\Big(\frac{R^*}{P}\Big)} + \frac{\mathrm{cov}\Big(\frac{R^*}{P},\ \frac{P^*}{P}\Big)}{\mathrm{VAR}\Big(\frac{R^*}{P}\Big)}Q(I).$$

The equilibrium domestic interest rate—derived in equation A-4—can alternatively be used in equation A-5, in which case the demand for foreign bonds will depend on the real interest differential and the covariance of the real interest differential with foreign interest rates.

The Generalized Case of N Foreign Bonds and L Domestic Investment Opportunities

In case of L domestic investment opportunities and N foreign bonds, there will be $1 + L + N$ first-order conditions: one for domestic bonds, one for each domestic investment opportunity, and one for each foreign bond. The N differ-

Table A-1. *Composition of Rolling Portfolios for Mexico, 1977–89*

Portfolio number	*Year*	*Month*	*U.S. dollar*	*DM*	*Yen*	*Amount to be borrowed as a percentage of export earnings*	*Mean square residual*	*Adjusted R^2*	*Durbin-Watson statistic*
1	1977	1	4.2936	0.8412	−6.135	−0.0018	0.04	0.24	1.97
2	1977	4	4.6652	0.9101	−6.575	−0.0016	0.05	0.23	2.04
3	1977	7	4.6664	0.9008	−6.567	−0.0016	0.05	0.23	2.11
4	1977	10	4.9198	0.9982	−6.918	−0.0015	0.05	0.22	2.12
5	1978	1	13.9847	2.9492	−17.934	−0.0005	0.04	0.21	2.09
6	1978	4	29.9681	7.2515	−36.220	0.0002	0.05	0.20	2.09
7	1978	7	0.8349	0.3283	−0.163	0.0044	0.05	0.17	2.10
8	1978	10	0.8376	0.3278	−0.165	0.0043	0.05	0.16	2.09
9	1979	1	0.8645	0.3724	−0.237	0.0042	0.04	0.14	1.91
10	1979	4	0.8470	0.3726	−0.220	0.0042	0.04	0.16	2.06
11	1979	7	1.4379	−0.0401	−0.398	0.0027	0.03	0.13	2.07
12	1979	10	0.8864	0.3306	−0.217	0.0040	0.03	0.14	2.06
13	1980	1	0.6550	0.2366	0.108	0.0045	0.03	0.13	1.99
14	1980	4	1.0901	−0.2329	0.143	0.0029	0.03	0.14	1.84
15	1980	7	1.0343	−0.1959	0.162	0.0031	0.02	0.16	1.97
16	1980	10	−0.8930	−0.3971	0.290	−0.0015	0.02	0.02	1.97
17	1981	1	1.0446	−0.0808	0.036	0.0078	0.02	0.10	1.91
18	1981	4	1.0866	−0.0834	−0.003	0.0078	0.02	0.15	1.66
19	1981	7	1.0078	−0.0326	0.025	0.0184	0.02	0.11	1.81
20	1981	10	1.0071	−0.0132	0.006	0.0373	0.03	0.06	1.92
21	1982	1	1.0094	−0.0114	0.002	0.0459	0.03	0.10	2.00
22	1982	4	1.0747	−0.0926	0.018	0.0074	0.03	0.06	1.84
23	1982	7	−0.5712	−0.0595	−0.369	−0.0105	0.03	0.06	1.97
24	1982	10	−0.6164	−0.0429	−0.341	−0.0139	0.04	0.06	1.97

25	1983	1	1.1022	−0.4414	−1.661	−0.0013	0.04	0.06	1.97
26	1983	4	1.3839	−0.5198	−1.864	−0.0012	0.04	0.07	1.96
27	1983	7	1.6296	−0.6995	−1.930	−0.0010	0.04	0.06	1.97
28	1983	10	1.6611	−0.7202	−1.941	−0.0010	0.04	0.06	1.92
29	1984	1	1.6979	−0.6765	−2.021	−0.0009	0.04	0.06	1.97
30	1984	4	1.8010	1.0433	−3.844	−0.0008	0.04	0.08	1.97
31	1984	7	1.7718	1.0413	−3.813	−0.0008	0.04	0.08	1.96
32	1984	10	1.8518	1.0763	−3.928	−0.0008	0.04	0.08	1.97
33	1985	1	1.8328	1.0542	−3.887	−0.0008	0.04	0.08	1.95
34	1985	4	3.1815	1.9637	−6.145	−0.0004	0.04	0.07	1.97
35	1985	7	1.4076	0.7766	−3.184	−0.0011	0.04	0.09	1.80
36	1985	10	4.3707	2.1911	−7.562	−0.0004	0.03	0.08	1.97
37	1986	1	17.6909	9.4037	−28.095	−0.0001	0.03	0.06	1.97
38	1986	4	0.4354	0.1885	−1.624	−0.0045	0.03	0.07	2.00
39	1986	7	0.4248	0.1910	−1.616	−0.0047	0.03	0.08	2.01
40	1986	10	2.4271	1.5531	−2.980	0.0006	0.02	0.09	1.94
41	1987	1	0.9899	0.0126	−0.003	0.0740	0.01	0.30	1.88
42	1987	4	1.0161	0.0128	−0.029	0.0754	0.01	0.33	1.82
43	1987	7	1.0080	0.0127	−0.021	0.0758	0.01	0.33	1.97
44	1987	10	1.0047	0.0127	−0.017	0.0752	0.01	0.32	1.95
45	1988	1	1.0116	0.0120	−0.024	0.0773	0.01	0.36	1.98
46	1988	4	1.0376	−0.0060	−0.032	0.0701	0.01	0.32	1.95
47	1988	7	1.0019	−0.0151	0.013	0.0718	0.01	0.32	1.95
48	1988	10	1.0050	−0.0192	0.014	0.0740	0.01	0.32	1.96
49	1989	1	1.0056	−0.0194	0.014	0.0745	0.02	0.32	1.93
50	1989	4	1.0051	−0.0199	0.015	0.0707	0.02	0.32	1.97
51	1989	7	1.0031	−0.0227	0.020	0.0591	0.01	0.31	1.92

Source: Author's calculations.

Table A-2. *Composition of Rolling Portfolios for Brazil, 1977–86*

Portfolio number	*Year*	*Month*	*U.S. dollar*	*DM*	*Yen*	*Amount to be borrowed as a percentage of export earnings*	*Mean square residual*	*Adjusted R^2*	*Durbin-Watson statistic*
1	1977	1	1.0178	0.0672	−0.0850	0.0762	1.31	0.21	2.05
2	1977	4	1.0267	0.0578	−0.0846	0.0744	1.37	0.20	2.11
3	1977	7	1.0074	0.0508	−0.0582	0.0899	1.30	0.23	2.13
4	1977	10	1.0079	0.0494	−0.0573	0.0932	1.24	0.26	2.12
5	1978	1	1.0588	0.0008	−0.0597	0.0626	1.04	0.14	1.90
6	1978	4	1.0809	−0.0247	−0.0562	0.0703	0.97	0.21	1.98
7	1978	7	1.0551	−0.0278	−0.0274	0.1306	0.88	0.27	1.97
8	1978	10	1.0480	−0.0191	−0.0289	0.1665	0.84	0.26	2.04
9	1979	1	1.0480	−0.0184	−0.0296	0.1571	0.81	0.25	1.96
10	1979	4	1.0388	−0.0226	−0.0162	0.1518	0.73	0.27	1.99
11	1979	7	1.0347	−0.0287	−0.0060	0.1378	0.74	0.28	2.01
12	1979	10	1.0582	−0.0505	−0.0076	0.1079	0.77	0.26	1.99
13	1980	1	1.0674	−0.0591	−0.0083	0.0877	0.77	0.30	1.86
14	1980	4	1.0925	−0.0854	−0.0072	0.0846	0.64	0.31	1.99
15	1980	7	1.1042	−0.0943	−0.0099	0.0820	0.63	0.32	2.02
16	1980	10	1.1144	−0.1047	−0.0097	0.0811	0.63	0.31	2.02
17	1981	1	1.1259	−0.1161	−0.0098	0.0814	0.60	0.32	2.03
18	1981	4	1.0628	−0.0531	−0.0097	0.0828	0.56	0.33	2.03
19	1981	7	1.0632	−0.0528	−0.0104	0.0787	0.60	0.29	1.99
20	1981	10	1.0693	−0.0626	−0.0067	0.0776	0.65	0.25	1.98

21	1982	1	1.1303	−0.1212	−0.0091	0.0744	1.81	0.12	1.91
22	1982	4	1.0617	−0.0521	−0.0095	0.0685	1.91	0.10	2.04
23	1982	7	1.1157	−0.1059	−0.0098	0.0682	1.91	0.11	2.09
24	1982	10	1.1387	−0.1286	−0.0101	0.0687	1.91	0.11	2.09
25	1983	1	1.1494	−0.1395	−0.0098	0.0687	1.95	0.10	2.07
26	1983	4	1.3053	−0.2942	−0.0111	0.0452	3.14	0.08	1.79
27	1983	7	1.4530	−0.5423	0.0893	0.0246	3.36	0.03	2.00
28	1983	10	2.0162	−1.2338	0.2176	0.0097	3.43	0.03	2.00
29	1984	1	−0.8810	−0.1323	0.0133	−0.1614	3.11	0.09	1.91
30	1984	4	−0.8935	−0.1175	0.0109	−0.1747	3.17	0.09	1.99
31	1984	7	−0.8876	−0.1233	0.0109	−0.1923	3.13	0.09	2.00
32	1984	10	−0.9029	−0.1446	0.0475	−0.1877	3.09	0.10	2.01
33	1985	1	−0.9721	−0.0697	0.0417	−0.1994	3.71	0.01	1.88
34	1985	4	−0.8184	−0.2692	0.0876	−0.1253	5.29	0.05	1.76
35	1985	7	−0.8112	−0.2730	0.0842	−0.1349	5.27	0.05	2.00
36	1985	10	−0.7615	−0.2835	0.0450	−0.1330	5.34	0.05	1.95
37	1986	1	−0.9655	−0.4183	0.3838	−0.1192	4.75	0.05	1.84
38	1986	4	−0.9275	−0.3841	0.3116	−0.1233	4.71	0.05	1.98
39	1986	7	−0.7465	−0.4378	0.1843	−0.1279	4.70	0.05	1.97

Source: Author's calculations.

ent first-order equations for the foreign bonds in equation A-2′ can be written in vector notation as one equation. Thus in equation A-6 V_{Rr} represents the vector of covariances of the real domestic costs with the real foreign effective costs, V_{rp^*} represents the covariance matrix of real domestic returns with the domestic unit value of export receipts, and V_{rr} represents the covariance matrix of real foreign returns.

$$\text{(A-6)} \qquad 0 = -U_1'1 + \beta E(U_2')\{E[(1/P)\mathbf{R}^*] - bV_{rp^*}Q(I) + bV_{rr}\mathbf{B} + bV_{Rr}1D\}.$$

Imposing that the domestic bond market is in zero net supply ($D = 0$), it can easily be shown that this equation can be solved for the optimal amount of foreign bonds as reported in equation 5 in the text.

The Speculative, Minimum Variance, and Price Hedges

When variables are lognormally distributed, we can split the speculative hedge and minimum variance hedge further and derive the price hedge. In the case of one foreign bond and one domestic investment opportunity we first combine equations A-2′ and A-3′, setting $D = 0$, to derive the following expression:

$$\text{(A-7)} \qquad E(U_2')\left[E\left(\frac{R^*}{P}\right) - E\left(\frac{R}{P}\right)\right] = b\text{cov}\left(\frac{P^*Q - R^*B}{P}, \frac{R^*}{P}\right) - b\text{cov}\left(\frac{P^*Q - R^*B}{P}, \frac{R}{P}\right).$$

Next we redefine variables in terms of percentage changes instead of second-period values. This implies that $E(R/P)$ can be written as $E(r) + E(1/p) + \text{cov}\ (r, 1/p)$, where lower case symbols are percentage changes. This also implies that the covariances can be rewritten. For instance:

$$\text{(A-8)} \qquad \text{cov}\left(\frac{P^*}{P}, \frac{R^*}{P}\right) \approx -\text{cov}(p^*, p) + \text{cov}(p, p) + \text{cov}(p^*, r^*) - \text{cov}(p, r^*).$$

Similar implications can be used for other covariances. Substituting these simplified expressions for the covariances in equation A-7, we see that some of the covariances will drop out. This allows us to rewrite equation A-7 and solve for the optimal amount of foreign bonds as a combination of a speculative, price, and minimum variance hedges, as in equation A-9:

$$\text{(A-9)} \qquad B = \frac{[E(r^*) - E(r)]E(U_2')}{-b\text{VAR}(r^*)} + \frac{\text{cov}(r^*, p^*)}{\text{VAR}(r^*)}Q(I) - \frac{\text{cov}(r^*, p)}{\text{VAR}(r^*)}\left[1 - \frac{E(U_2')}{b}\right].$$

For an alternative derivation, see Claessens (1988).

References

The word "processed" describes informally reproduced works that may not be commonly available through library systems.

Adler, Michael, and Bernard Dumas. 1980. "The Exposure of Long-Term Foreign Currency Bonds." *Journal of Financial and Quantitative Analysis* 15: 971–94.

Breeden, Douglas. 1979. "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities." *Journal of Financial Economics* 7: 265–96.

Claessens, Stijn. 1988. "The Optimal Currency Composition of External Debt." PRE Working Paper 14, World Bank, International Economics Department, Washington, D.C. Processed.

Dornbusch, Rudiger. 1985. "Policy and Performance Links between LDC Debtors and Industrial Nations." *Brookings Papers on Economic Activity* 2: 303–56.

———. 1987. "Exchange Rates and Prices." *American Economic Review* 77: 93–106.

Eaton, Jonathan. 1990. "Debt Relief and the International Enforcement of Contracts." *Journal of Economic Perspectives* 4(1):43–56.

Fraga, Arminio. 1986. "The Exchange Rate Premium and the Micro-Foundations of Asset Markets Models of the Exchange Rate." *Journal of International Economics* 20: 179–85.

Gemmill, Gordon. 1985. "Optimal Hedging on Futures Markets for Commodity-Exporting Nations." *European Economic Review* 27: 243–61.

Giovannini, Alberto. 1986. "The Macroeconomics of Exchange Rate Changes and Price-Level Interactions: Empirical Estimates for West Germany." Working paper, Columbia University, Graduate School of Business, New York. Processed.

Healy, James. 1981. "A Simple Regression Technique for the Optimal Diversification of Foreign Exchange Reserves." International Monetary Fund, Research Department, Washington, D.C. Processed.

IMF (International Monetary Fund). Various years. *International Financial Statistics*. Washington, D.C.

Kahn, R. B. 1988. "Currency Diversification of Developing-Country External Debt with Liquidity Constraints." Federal Reserve System, Division of Research and Statistics, Washington, D.C. Processed.

Kroner, Kenneth, and Stijn Claessens. 1991. "Optimal Dynamic Hedging Portfolios and the Currency Composition of External Debt." *Journal of Money and Finance* 10: 131–48.

Krugman, Paul. 1981. "Consumption Preferences, Asset Demands, and Distribution Effects in International Financial Markets." NBER working paper 651. National Bureau of Economic Research, Cambridge Mass. Processed.

Lessard, Donald, and John Williamson. 1985. *Financial Intermediation beyond the Debt Crisis*. Policy Analysis In International Economics 12. Washington, D.C.: Institute for International Economics.

Macedo, J. B. de. 1982. "Portfolio Diversification across Currencies." In R. N. Cooper and others, eds., *The International Monetary System under Flexible Exchange Rates*. Cambridge, Mass.: Ballinger.

———. 1983. "Optimal Currency Diversification for a Class of Risk-Averse International Investors." *Journal of Economic Dynamics and Control* 5: 173–85.

Merton, R. C. 1971. "Optimum Consumption and Portfolio Rules in a Continu Time Model." *Journal of Economic Theory* 3: 373–413.

Oxelheim, Lars, and Clas Wihlborg. 1987. "Exchange-Rate-Related Exposure in a Macroeconomic Perspective." In Sarkis Khoury and Alo Ghosh, eds., *Recent Developments in International Banking and Finance*. Lexington, Mass.: Lexington.

Stulz, Rene. 1981. "A Model of International Asset Pricing." *Journal of Financial Economics* 9: 383–406.

———. 1984. "Optimal Hedging Policies." *Journal of Financial and Quantitative Analysis* 19: 127–40.

Svensson, L. E. O. 1989. "Portfolio Choice and Asset Pricing with Nontraded Assets." NBER working paper 2774. National Bureau of Economic Research, Cambridge, Mass. Processed.

Varangis, Panayotis, and Ron Duncan. 1987. "Exchange Rates Pass-through in Primary Commodity Prices." World Bank, International Economics Department, Washington, D.C. Processed.

World Bank. 1987. *The World Bank Annual Report*. Washington, D.C.

Household Saving in Developing Countries: First Cross-Country Evidence

Klaus Schmidt-Hebbel, Steven B. Webb, and Giancarlo Corsetti

Although most studies have relied on domestic or private sector saving data, this article uses household data available from the U.N. System of National Accounts for a sample of 10 countries. Household saving functions are estimated using combined time-series and cross-country observations in order to test households' responses to income and growth, rates of return, monetary wealth, foreign saving, and demographic variables. The results show that income and wealth variables affect saving strongly and in ways consistent with standard theories. Inflation and the interest rate do not show clear effects on saving, which is also consistent with their theoretical ambiguity. Foreign saving and monetary assets have strong negative effects on household saving, which suggests the importance of liquidity constraints and monetary wealth in developing countries.

Households are responsible for a substantial part of the saving in both industrial and developing countries. Accordingly, most economic models treat the motivation for saving from the household's perspective. Because of the shortage of household data, however, most empirical work on saving in developing countries has used only aggregate saving data, so that the results may not be useful for predicting household behavior.

This study tests several hypotheses about saving behavior using panel data from the U.N. System of National Accounts. This is the first study to use cross-country variation to estimate household saving behavior in developing countries and has the advantage that the data are defined consistently across countries. The study tests how household saving in developing countries responds to income and growth, rates of return, monetary wealth, foreign saving, and demographic variables.

Section I reviews the evidence in the empirical literature on the main determinants of saving in developing countries. Section II presents an empirical frame-

Klaus Schmidt-Hebbel and Steven B. Webb are with the Country Economics Department at the World Bank; Giancarlo Corsetti was with the department when this article was written. They are indebted to Ricardo Caballero, Vittorio Corbo, Stanley Fischer, Mark Gersovitz, and two anonymous referees for helpful comments and suggestions and thank Heidi Zia for assistance with some of the research for this article.

work for analyzing household saving in developing countries, discusses the data set and estimation methods, and presents results. Section III concludes with a discussion of policy implications.

I. Determinants of Saving in the Empirical Literature

A number of studies have estimated saving behavior in developing countries. (For general surveys, see Mikesell and Zinser 1973, Gersovitz 1988, and Deaton 1989.) Typically, studies use cross-section, time-series data on national saving rates; see for instance Collins (1989); Fry (1978, 1980, 1988); Giovannini (1983, 1985); and Gupta (1987). The advantage of using national saving rates is that more years of data are available for more countries. Even when private or household data are available, however, as in Collins (1989), they may have limited use because of lack of comparability across countries. The use of national saving is rationalized with the argument that private saving is a large and typically predominant part of total saving.

To extrapolate from total saving behavior to household saving behavior, however, requires the assumption that the latter is a perfect substitute for both private corporate saving and public sector saving. Assuming substitutability between private and public saving requires that Ricardian equivalence holds. If income and saving data for just the household sector are used, then household saving decisions are not necessarily assumed to offset saving decisions made elsewhere in the economy, unless households respond to determinants such as disposable income, public saving, or private wealth in a way consistent with the Ricardian hypothesis.

A few studies have used private sector saving or household consumption data, but none of the studies with aggregate data from developing countries has focused on household saving. Because data of this type are only gradually becoming available, the data sets vary widely from study to study. Rossi (1988) uses a cross-section, time-series data set for 49 countries covering 10 years. Saving is implicit because per capita private consumption is estimated as a function of per capita private income, among other things. Private income includes the profits of public as well as private enterprises, so that the implicit saving includes saving by public sector enterprises. Lahiri (1988) uses time-series data for private consumption to run separate regressions for 8 Asian countries with about 20 years of data for each. Several older studies, such as Singh (1975) and Williamson (1968), also used private saving data, but they investigated Keynesian theories of saving and overlooked many of the issues currently of interest. Despite differences in the data, most studies using private sector data produce consistent results on most of the major issues. The main determinants of private saving or consumption considered in the literature fall into five groups: income and wealth, public saving, rates of return, foreign saving, and demographic variables.

As discussed below, there is a broad consensus that faster economic growth

and high incomes contribute to higher saving rates. Evidence on the pervasiveness of liquidity constraints suggests that an increase in public saving is not neutralized by an offsetting decline in private saving. There is still controversy and lack of sufficient evidence about the effects of monetary asset holdings, foreign capital inflows, the real interest rate, and inflation, as well as demographic variables.

Income and Wealth

Income and wealth are the basic determinants of consumption. Most studies use per capita income levels or growth rates and total or financial wealth as the relevant income and wealth variables.

The level of per capita income is hypothesized to have a positive effect on the saving rate because richer people can afford the luxury of saving to assure their future consumption. The poor are more likely to be at their biological or social minimum level of current consumption. This does not mean zero saving by the poor in all years, for they will attempt to cushion themselves against fluctuations in current income. But they will have relatively smaller cushions and will more frequently find themselves with zero wealth and no opportunity to borrow in order to sustain consumption when income is low (Deaton 1989; Zeldes 1989). All of the empirical studies mentioned earlier find a strong positive effect—measured in various ways—of the current income level on the saving rate.

Saving may also depend on income fluctuations. Simple consumption-smoothing models predict that temporary fluctuations in income should primarily affect saving. If households are not credit constrained and the temporary fluctuation does not change much permanent income, consumption would respond only marginally to temporary income fluctuations. Households, particularly in developing countries, tend to be credit constrained, however, which would imply that consumption would respond significantly to temporary income changes. Campbell and Deaton (1989) also argue that, at least in industrial countries, a household's perception of its permanent income is strongly affected by current shocks, with no evident distinction made between current and permanent income flows. This implies that households will consume out of current shocks, saving much less out of shocks than predicted by consumption-smoothing models. Most empirical studies of developing countries have not looked closely at the effect of income fluctuations on saving. An exception is Gupta (1987), which consistently finds that saving responds significantly and positively to temporary income shocks.

The growth rate of income is also typically included in recent saving studies for developing countries, but its effect cannot be signed *a priori*. Intertemporal models of consumption or saving (that is, permanent income or life-cycle models) predict that an increase in the growth rate of an individual household's income would lower its saving rate; households would save less now if they expected higher incomes in the future, which would allow both higher consumption and higher saving in the future. Faster growth of average per capita (or

average household) income could also have positive effects on saving. More rapid income growth would raise the average household saving rate if the growth were concentrated in households with high saving rates, such as rich or middle-age households. Collins (1989) develops these concepts in a simple model. A positive coefficient for the growth rate could also reflect the slowness with which people change their consumption habits or indicate that people have regressive expectations about their level of income. The studies that investigate real growth of gross domestic product (GDP) as a determinant of saving—such as Collins (1989); Fry (1978, 1980); Giovannini (1983, 1985); Mason (1987, 1988); and Mason and others (1986)—find positive and usually significant effects on the saving rate.[1]

Wealth is usually a key determinant of consumption or saving in theoretical models of intertemporal optimization. Of course, permanent income can be viewed as the stream of income from total wealth, but a narrower definition of wealth would be the assets that can be exchanged for current consumption. Theory unambiguously predicts that greater wealth would reduce saving out of current income. Since most concepts of wealth are not directly observable, wealth has not been used in most empirical studies of saving in developing countries. Schmidt-Hebbel (1987) uses five alternative measures of total wealth in an empirical intertemporal consumption model for Chile, based on different assumptions regarding expectation formation. Behrman and Sussangkarn (1989) use household-level data on wealth and saving in Thailand. In both studies, wealth has a strong negative effect on saving.

Monetary or financial assets also lessen a household's dependence on current income sources when income declines temporarily, because consumers can draw on the assets to maintain their consumption levels. Hence, holding a higher stock of assets allows a household to maintain a higher consumption rate on average, thus depressing the saving rate. In addition, monetary asset holdings are an important component of total consumer wealth. This implies that monetary holdings have a second, negative influence on saving rates; previous studies have not tested for this.

Government Deficits and Government Saving

Fiscal stabilization programs aim to lower public sector deficits or raise public saving, which is equal to public investment minus the deficit. Changes in government deficits or government saving may change households' perceived permanent income and therefore affect private saving rates. A voluminous empirical literature has been devoted to the Ricardian equivalence proposition as reformu-

1. There might be some simultaneity bias—both high saving and high growth reflecting the effect of good investment opportunities, which are not fully reflected in the real deposit interest rates. Also, with the public sector included in aggregate saving, one would expect revenues to adjust more automatically to income increases than current expenditures. Furthermore, government investment (which counts as saving) might be driving both faster growth and higher aggregate saving.

lated by Barro (1974), which states that public debt issues are macro-economically indistinguishable from tax increases, and hence a change in public saving should be offset by an equal and opposite change in private saving.

For Ricardian equivalence to hold, the following conditions must be satisfied jointly: no liquidity constraints; equal discount rates for the public and private sectors; and certainty of future income, tax, and public expenditure flows. For a further discussion of these conditions and surveys of empirical studies, see Hayashi (1985), Hubbard and Judd (1986), Bernheim (1987), and Leiderman and Blejer (1988). Ricardian equivalence has been widely rejected in empirical studies of industrial countries. The pervasiveness of borrowing constraints is most often cited as the main reason for its rejection. For a sample of developing countries, Haque and Montiel (1989) also find that borrowing constraints are the main cause of the deviations from Ricardian equivalence. They also test for differences between public and private discount rates but do not find that such differences explain the deviations from Ricardian equivalence. Credit constraints, proxied by current income levels or financial asset holdings, are also major determinants of private saving in the cross-developing-country study by Rossi (1988). Corbo and Schmidt-Hebbel (1991) explicitly test the Ricardian hypotheses by regressing private consumption on permanent government saving in 13 developing countries. They conclude that the Ricardian hypothesis does not explain consumption behavior. Because it implies that private and public saving are not perfect substitutes, empirical rejection of Ricardian equivalence is a powerful argument in favor of using household saving data instead of aggregate saving data.

Rates of Return

Because the income and substitution effects of higher interest rates work in opposite directions, the effect of rates of return on saving cannot be predicted. In addition to the well-known income and substitution effects, rates of return also affect saving through a wealth effect. A higher real interest rate reduces the present value of future income streams from human capital or fixed-interest financial assets. Consumption is therefore depressed even if the substitution and income effects cancel each other. Economists have heatedly debated the interpretation of the empirical evidence. Based on noneconometric country studies, McKinnon (1973) and Shaw (1973) argued that the rate of return on saving, as measured by the real interest rate, would have a positive effect on saving rates. This view was also held by Balassa (1990) in his survey on the issue. Fry (1978, 1980) found statistical evidence to support this contention, although he conceded that the magnitude of the effect is small (Fry 1988). Thus, only large changes in real interest rates would be economically important. Giovannini (1983, 1985) revisited Fry's earlier work and found that two observations (on the Republic of Korea in 1967 and 1968) accounted for the entire result. With an expanded data set, Giovannini found that the interest rate did not signifi-

cantly affect saving. Both Fry and Giovannini used aggregate data, which is especially problematic for testing the effects of interest rates. Giovannini (1985) implicitly recognizes this; he points out that much of the increase of Korean saving in the late 1960s resulted from an increase in government saving. Gupta (1987) finds some support for a positive effect of interest rates on saving in Asia, but not in Latin America. Schmidt-Hebbel (1987) and Arrau (1989) estimate intertemporal consumption substitution elasticities for Southern Cone countries and find values close to 1.0. With an elasticity of 1.0 and for given wealth stocks, consumption is insensitive to the interest rate because the substitution effect is canceled by the income effect.

Even if higher real interest rates are unlikely to raise private saving and hence total private wealth, they can substantially alter the portfolio composition of private wealth. Negative real interest rates on deposits (resulting from high inflation), for example, cause substitution into real assets, especially consumer durables, and into foreign currency assets through capital flight. Both higher purchases of durables and capital flight tend to reduce private saving as measured by official national accounts, without affecting private saving defined as total private wealth accumulation.

Inflation may affect saving independently from its effect through the real interest rate. High inflation often contributes to stagnation in output or outright recession, which will be reflected in the income variable. Higher inflation also increases economic instability and uncertainty about future rates of return on real assets and income levels. Theoretically, the effect of inflation on private saving is ambiguous, as uncertainty about future asset values could either discourage saving because of the substitution effect of the lower real rate of return or encourage saving for precautionary motives. Similarly, the increase in uncertainty of future income streams affects private saving ambiguously, depending on the form of the underlying utility function. A comprehensive treatment of the effects of different sources of uncertainty on saving can be found in Gersovitz (1988).

Gupta (1987) and Lahiri (1988) include the expected and unexpected components of the inflation rate as separate determinants of saving, and Gupta also includes the nominal interest rate. Gupta's results differ sharply by region. In Asia both expected and unexpected inflation have positive and significant coefficients. In Latin America neither coefficient was significant with the preferred estimation technique. In Lahiri's sample, comprised only of Asian countries, the signs on both inflation variables are mixed for the eight separate country regressions. Therefore, there is no clear empirical evidence that inflation generally affects saving in developing countries.

Foreign Saving

If access to foreign borrowing at international interest rates is unlimited, foreign saving passively fills the gap between domestic investment and national saving. In this case, foreign saving (or borrowing from abroad) is simply the

result of national saving decisions and not one of its determinants. If foreign borrowing is rationed, however, either by the lenders or by government regulation in developing countries, then domestic savers (and investors) are constrained in their intertemporal choices by the amount of available foreign finance, and foreign saving becomes a determinant of domestic saving.

During most of the post–World War II period, developing countries have not had unrestricted access to private commercial lending, a partial exception being the period from 1976 to 1981, before the debt crisis erupted. Even during that short time, governments in developing countries restricted access to foreign borrowing. When restrictions were partially lifted, governments and corporations typically made use of their freer access to foreign borrowing to increase investment, but households were the last and least likely sector to have unlimited access to sources of foreign loans. Hence, even during 1976–81, households were effectively constrained by foreign lending: for given income levels, a fraction of total foreign lending finances higher private consumption, hence reducing household saving.

A number of empirical studies have included foreign saving as a determinant of domestic saving rates. Fry (1978, 1980) and Giovannini (1985) find the effect of foreign saving is significant and negative, although its size is small. With a noneconometric analysis, Chenery and Strout (1966) also find a negative initial impact of capital inflows on domestic saving, although the secondary effects on capacity growth tend to increase saving. Giovannini (1983) finds coefficients on foreign saving to have mixed signs and to be insignificant. Gupta (1987) finds positive coefficients on foreign saving, which are significant for Latin America but not for Asia. The results of these studies seem to depend on the sample and model specification. All of the studies with foreign saving as an explanatory variable used total saving as the dependent variable, so the results may reflect the extent to which capital inflows went straight to public and corporate investment, which appears as an increase in saving.

Demographic Variables

The life-cycle models of saving imply that demographic variables should affect saving rates, and demographic influences on saving have been widely researched (Collins 1989; Leff 1969; Mason 1987, 1988; Mason and others 1986; Rossi 1989; Webb and Zia 1990; see Hammer 1986 for a survey). The dependency ratio—those under age 15 or over 65 as a share of total population—is the most commonly used explanatory variable. In the life-cycle model, older people work less and consume out of their savings. Households with more children at home may also save less because saving for retirement would be deferred until the children left home, which would raise the per capita income of the household, or because parents would expect old-age support from their children. Thus, one would expect saving rates to depend negatively on the dependency ratio.

Early work on the topic, especially that of Leff (1969), found that the dependency ratio had a strong negative effect on saving. Subsequent studies challenged

the robustness of this result and have examined both the theory and the measurement of demographic variables more carefully (Mason and others 1986; Mason 1987, 1988). The results seem to depend on the data used and on the other explanatory variables included. The models of Mason and others (1986) and Collins (1989), applied to cross-country samples of Asian economies, add to the dependency ratio and the growth rate of per capita income (among other variables) an interaction term between the two. Although Mason and others find that higher population dependency lowers saving unambiguously, Collins's results show an ambiguous effect of the dependency ratio, with a negative (positive) influence on saving in countries with high (low) growth rates.

II. New Evidence on Household Saving

To test for the determinants of saving with household data, a behavioral function for household saving was estimated. The function incorporates variables to address the major unresolved issues in the literature. The saving rate (the ratio of household saving to private disposable income), rather than the absolute level of saving, is used as the dependent variable for three reasons. First, there is no adequate deflator for saving that can be used to obtain constant-price saving series. Second, by using ratios instead of levels, cross-country comparisons can be made without having to choose appropriate exchange rates. Third, saving rates tend to be stationary, whereas absolute saving flows grow over time, so that, by using rates, spurious correlation with time-trended explanation variables can be minimized.

The specification for the household saving rate is consistent with the consumption hypotheses reviewed in the previous section but is not derived from an explicit optimizing framework. The reason for the latter is the difficulty of deriving a closed-form saving equation with the host of economic and structural determinants considered in this study. Hence the following general saving function is implemented in a linear form:

$$(1)\qquad \frac{S}{I}=\frac{S}{I}\left[\underset{(+)}{\text{LITP}},\ \underset{(+)}{\text{GITP}},\ \underset{(+)}{(\text{LIP}-\text{LITP})},\ \underset{(?)}{\frac{\text{HT}}{I}},\ \underset{(?)}{R},\ \underset{(?)}{\text{INF}},\ \underset{(-)}{\frac{\text{MQM}}{I^{*}}},\ \underset{(-)}{\frac{\text{FS}}{I}},\ \underset{(?)}{\text{DEP}},\ \underset{(?)}{\text{URB}}\right]$$

where S is household saving, I is household disposable income, LITP is the natural logarithm of trend per capita household disposable income, GITP is the growth rate of trend per capita household disposable income, LIP is the natural logarithm of per capita household disposable income, HT is transfers to households, R is the real interest rate, INF is the inflation rate, MQM is money plus quasi-money at the end of the previous period, I^* is the average of household disposable income in the current and previous years, FS is foreign saving, DEP is the dependency ratio, and URB is the urbanization rate. Signs below the variables indicate the expected coefficient signs, given the discussion of the preceding section. Equation 1 generalizes the saving functions estimated for developing countries in the recent empirical literature by including explanatory variables

each of the five main groups: four income determinants, two rates of n, broad money, foreign saving, and two demographic variables.

e first three income variables are the natural logarithm of trend per capita te disposable income, its growth rate, and the deviation of log current from end per capita private disposable income. Unlike previous studies, both the l and the deviation of current from trend per capita income are introduced ving determinants in order to discriminate between the permanent and al influences of income on the saving rate. Because the sum of log trend ne and the deviation of log current from log trend income equals log current ne, we can test the proposition, implicitly assumed in some earlier studies, the coefficients on the two components of income are the same. Growth of apita trend income enters as a separate, third determinant. Finally, in order st whether saving out of transfers differs from saving out of other income, fers to households is included. Since transfers are already counted as part of household income, the coefficient on transfers would be zero if consump- out of that income were the same as that out of other income. A significant tive coefficient would indicate that households saved less out of transfers out of other income.

mestic real interest and inflation rates, both with ambiguous signs, could t intertemporal consumption and portfolio composition decisions, with equences for household saving. Monetary wealth (which reflects both con- r wealth and liquidity constraints) and foreign saving should depress house- saving rates. Finally, the two demographic variables have ambiguous signs. ugh the dependency ratio should reduce saving due to life-cycle patterns, t influence on saving seems to depend on its measurement and the inclusion her variables. The urbanization variable was added to control for the effect fferences in the measurement of urban and rural saving, as well as for tural differences in the underlying saving behavior of urban and rural eholds.

estimating equation 1, simultaneity biases could arise from the potential geneity of some explanatory variables. For this study the potentially most rtant simultaneity is between household saving and the real interest rate, h will be addressed by instrumenting the latter variable. The responsiveness reign saving to household saving appears to be minimal in the sample of loping countries used, for the reasons discussed in section I. Also, house- saving does not appear to have important contemporaneous effects on ehold income, because even the effects on GDP, from which household ne derives, are dominated by other influences. Hence neither foreign saving ousehold income will be instrumented.

Data

e data set is especially well suited for investigating household saving behav- t is based on household saving and disposable income series for 10 econ- s for which at least 7 and as many as 13 consecutive annual observations available during 1970–85. The economies are Botswana, Colombia, Ecu-

ador, Honduras, Republic of Korea, Philippines, Paraguay, Thailand, South Africa, and Taiwan. Except for Taiwan, the data come from the U.N. System of National Accounts, which breaks down income and consumption into general government, corporate, and household sectors. From these, household disposable income, household saving, and transfers from the general government to households were calculated. The data for Taiwan come from the *Statistical Yearbook of the Republic of China*, compiled by the Taiwanese government using the U.N. method. All household data (saving, disposable income, and transfers) are aggregate data obtained from national accounts sources, not microeconomic data from household surveys. Interest rates come from a data set developed by the Country Economics Department of the World Bank. The remaining data (inflation, urban population share, dependency ratio, current account balance, and money balances) are from the International Monetary Fund's *International Financial Statistics* and *Government Financial Statistics* and the World Bank Economic and Social Database.

Household disposable income includes all current receipts by households, less taxes and social security contributions. One potential problem is that excluding the retained earnings of (private) corporations owned by households may distort the picture of household decisionmaking. The household sector does include all agricultural firms and firms in the informal nonagriculture sector. Excluding the income of private corporations will not affect the results as long as most of the variation of household income and saving is accounted for by households that would not consider the corporate income and saving as part of their own budget and would not make household saving decisions to offset corporate sector decisions. Another potential problem is that in the System of National Accounts net saving is defined as the difference between current receipts and current outlays and is therefore a balancing item. Hence measurement errors in receipts and current outlays are reflected in measured saving. As long as these errors are systematic and stable over time for each country, they are completely accounted for by the use of country-specific intercepts in the empirical model introduced below.

To calculate the three income variables, the log of household disposable income was regressed on a time trend using five-year overlapping data series, up to and including the current year. The estimated value for the current year gives the trend value of current income, the coefficient for time is the trend rate of growth, and the deviation from the estimated value in the current year is the temporary fluctuation in income. This simple deterministic time trend method was used instead of more sophisticated time-series procedures because of lack of data. Although one should view the results with caution, this method avoids some of the pitfalls of using simple moving averages. To obtain results for the first sample observations for each country, household income for the four years before the start of the sample was estimated using the growth rate of GDP in each year. Transfer income includes social security, unemployment insurance, and transfers from abroad; it does not include indirect transfers such as farm price supports or subsidies that permit food to be sold at below-market prices.

The nominal interest rate is the annual rate on three-month time deposits. The inflation rate is the annual (December-to-December) change in the log of the household consumption deflator. The sample includes countries with low to medium inflation, with annual average inflation rates ranging from 7 percent in Honduras to 23 percent in Colombia. Household income, which includes net nominal interest income, is therefore only slightly distorted by including the inflationary component of interest income. Data to correct for this bias are not available for the sample of countries used. The real interest rate is calculated as: $R = [(1 + \text{NOI})/(1 + \text{INF})] - 1$, where R and INF were introduced in equation 1 above and NOI is the nominal interest rate. Money and quasi-money holdings at the end of the previous period were used as a measure of liquid wealth available for consumption in the current period. To obtain its value as a share of income, the sum of money and quasi-money holdings was divided by the geometric average of nominal disposable income in the current and previous years, which is an adequate measure for low- to medium-inflation countries. Foreign saving is the current account balance, again as a share of household disposable income.

The two demographic variables, the dependency and urbanization rates, change little for each country during the observation period and in many cases are not known on an annual basis. Because both variables vary widely across countries but little over time, collinearity between country dummies and these variables is unavoidable. The dependency ratio is the population below 15 years and above 65 years as a percentage of total population. The urbanization rate is the share of population in cities, originally put together by the United Nations; the population cutoff for an area to be designated as a city varies from country to country.

Estimation Methods

The base specification is a fixed effect model of the form:

$$Y_{it} = \beta_0 + \sum_{i=2}^{N} \gamma_i W_{it} + \sum_{k=1}^{K} \beta_k X_{kit} + \epsilon_{it} \tag{2}$$

where the subscript i refers to countries, t refers to time, and k refers to independent variables. Y_{it} is the vector of saving rates, W_{it} is the vector of country dummy variables (such that W_{it} is equal to 1 for the ith country, and 0 otherwise), X_{kit} is the matrix of independent variables, and ϵ_{it} is the vector of errors, which are assumed to satisfy the assumptions of the classical normal linear model.

This basic model incorporates fixed country-specific effects in the intercept term. In fixed-effects models, the empirical results are conditional on the particular sample used in the estimation. Alternative estimators, such as the error-components (or random-effect) model, treat the available observations as a random sample from a population. Given the small number of countries included here and the marked differences in their economic features, the fixed-effect estimator seems to be an appropriate choice.

Table 1. *Determinants of Household Saving, with Household Sector Saving as a Percentage of Household Disposable Income as Dependent Variable*

Estimation technique	Independent variable										Adjusted R^2
	Trend income (log)	Income growth rate (five-year average)	Deviation of income from trend (log)	Transfers to households (ratio to income)	Real interest rate	Inflation rate	Beginning of period money and quasi-money (ratio to income)	Foreign saving (ratio to income)	Dependency ratio	Urban population ratio	
1. OLS[a]	0.03 (1.4)	0.75 (6.2)	0.12 (0.8)	−0.32 (−1.7)	−0.10 (−0.8)	−0.06 (−0.6)	−0.03 (−1.0)	−0.23 (−4.3)	−0.48 (−3.2)	−0.001 (−2.1)	0.646
2. Fixed effect	0.26 (5.4)	0.54 (4.2)	0.30 (2.5)	−0.40 (−2.1)	−0.08 (−0.7)	−0.14 (−1.5)	−0.19 (−3.1)	−0.14 (−2.8)	0.83 (2.5)	0.0001 (0.1)	0.811
3. Random effect[b]	0.04 (1.6)	0.70 (5.8)	0.19 (1.3)	−0.38 (−1.9)	−0.13 (−1.1)	−0.13 (−1.2)	−0.06 (−1.7)	−0.18 (−3.4)	−0.54 (−3.1)	−0.001 (−1.6)	0.703
4. Fixed effect with instrumental variables	0.25 (4.6)	0.51 (3.5)	0.27 (2.0)	−0.42 (−2.1)	−0.19 (−0.7)	−0.23 (−1.0)	−0.18 (−3.0)	−0.12 (−2.2)	0.75 (2.0)	0.0004 (0.2)	0.809
5. Fixed effect with instrumental variables[c]	0.22 (3.1)	0.56 (3.9)	0.26 (1.8)	−0.19 (−0.7)	−0.15 (−0.5)	n.a. n.a.	−0.16 (−2.3)	−0.14 (−2.7)	0.56 (1.0)	−0.0004 (−0.2)	0.765
6. Fixed effect[d]	0.26 (5.7)	0.56 (4.4)	0.31 (2.8)	−0.38 (−2.0)	n.a. n.a.	−0.08 (−1.9)	−0.18 (−3.2)	−0.14 (−3.1)	0.87 (2.7)	−0.0001 (−0.04)	0.812

n.a. Not applicable.

Note: t-statistics are in parentheses.

a. Breusch-Pagan test for the null hypothesis of the absence of individual effects in the errors: $\chi_1^2 = 15.99$ (P-value = 0.00006).

b. Hausman specification test for the null hypothesis of random-effect estimators equal to fixed-effect estimators: $\chi_1^2 = 28.4$ (P-value = 0.002).

c. Inflation rate is omitted variable.

d. Real interest rate is omitted variable.

Source: Authors' calculations.

This choice has also been verified by the usual set of specification tests. (For a detailed description of these tests, see Kmenta 1986 and Hsiao 1986.) The test results are reported in table 1. The first of these is the Breusch-Pagan test for the presence of both individual (cross-section) and time effects in the residuals of a simple ordinary least squares (OLS) regression on pooled data. The OLS specification was rejected, which implies that either the fixed effects model or the random effects model is superior. The second test applied is the Hausman specification test, which compares fixed-effect and random-effect estimators. Random-effect estimators are efficient but are consistent only in the absence of correlation between the included regressors and the errors. Fixed-effect estimators are not efficient but are still consistent even when the regressors and errors are correlated. The random effects estimates were found to be inconsistent, implying that a fixed-effects specification is superior. Finally, regression *t*-tests for the presence of trend- and time-specific effects showed such effects were never significant in any of the models specified here.

An important caveat to the results below is that the number of cross-sectional units considered in the estimation is small, which is problematic because the properties of panel data estimators are based on asymptotic results for a large number of cross-sectional units. This problem could not be avoided because the construction of a data set useful for testing household saving behavior limits the sample to only those countries for which disaggregated data on saving are available over a sufficiently long period. Its implication is that the empirical results reported below could change in the future when more household data at the country level become available.

Results

Table 1 summarizes the regression results with different estimation methods and with different sets of regressors. Four estimation techniques were used: OLS, country-specific fixed effects, country-specific random effects, and fixed effects instrumental variables to instrument the real interest rate. It was expected that the country-specific, fixed-effects method—with or without instrumental variables—would be best, and the results in table 1 confirm this.

Instrumental variable estimation corrects for the possible simultaneity bias stemming from the interaction between saving and domestic real interest rates. LIBOR, corrected for world inflation, was used as the instrument for the domestic real interest rate. Other sets of instrumental variables, such as the lagged independent variables, were also tried but did not improve upon the reported results. Regression 4 in table 1—fixed-effects estimation with instrumentalization of the real interest rate—thus constitutes the primary set of results.

The estimated regressions fit the data well, as can be seen in the adjusted R^2 coefficients. Starting at 0.65 for the OLS estimation, the adjusted R^2 increases to 0.81 under fixed-effects estimation. This overall fit compares favorably with other panel data estimations of saving functions in developing countries, such as Corbo and Schmidt-Hebbel (1991), which explain roughly half of the variation of the dependent variable.

The income variables all have a positive effect on saving rates, which accords with most of the previous studies for developing countries. The growth rate of trend per capita disposable income has a strong effect on the household saving rate in the sample countries: a 1 percentage point increase in per capita income growth raises the household saving rate by about 0.5 percentage points. An increase in the growth of per capita income seems the best way to raise the private saving rate, which suggests the presence of a virtuous circle between growth and saving.

The deviation of income from its trend level has a positive influence on the saving rate. Its coefficient is in the neighborhood of 0.30, however, which is much lower than that predicted by the permanent income theory. Given the saving function in equation 1, it is easy to show that this coefficient is equal to the difference between the average and marginal propensities to consume out of current (transitory) income. Only if the estimated value of this coefficient were close to 0.75, which is the average propensity to consume out of current income in the sample, would the marginal propensity be close to zero, as predicted by the permanent income hypothesis. This effect of transitory income on consumption reflects the importance of borrowing constraints faced by households or their use of current income in estimating permanent income levels.

The dominance of transitory over trend (permanent) income in determining household consumption is confirmed by the coefficient of the log of the level of per capita trend disposable income. Although it is highly significant, one cannot reject the hypothesis that it is identical to the estimated coefficient of the deviation of current from trend income. According to the permanent income hypothesis, the marginal propensity to save out of permanent income should be close to −1. For the saving equation specified here, this would require that the coefficient of trend income should be equal to the difference between the coefficient of the deviation of current from trend income and 1—a far cry from the actual results. This implies that, for the present sample, trend income could be omitted as a determinant of household saving, using instead just current income rather than decomposing income into its trend and fluctuation. With regard to the last income variable, the coefficient on transfers is negative and usually large and significant, thus indicating that households consume more out of transfers on average than out of other sources of disposable income.

The domestic real interest rate has a small, mostly negative, and not significant influence on household saving rates. This result, confirming many recent studies that show a negligible role of interest rates in determining consumption or saving, reflects either that income and substitution effects offset each other or that liquidity constraints weaken the effects of intertemporal relative prices on intertemporal consumption choices. The important role of liquidity constraints in our sample is corroborated by the significance of current income, monetary wealth, and foreign saving in affecting household consumption and saving.

Inflation has a negative but statistically not significant effect on saving. Because of collinearity between the real interest rate and the inflation rate, these

variables were entered alternately, as reported in regressions 5 and 6. The inflation rate is significant at about the 10 percent level when the real interest rate is excluded, and its coefficient is negative. Omitting inflation, however, does not make the effect of the real interest rate significant. In other words, reducing inflation seems to encourage household saving, but raising the deposit rate relative to inflation has no discernible effect on saving.

Monetary assets play a dual role here: first, they constitute a stock variable signaling the extent of domestic liquidity constraints; second, they are related to household financial wealth. For both reasons monetary stocks were expected to have a negative influence on saving. The results show a negative and significant coefficient for the ratio of money to income in the fixed-effect estimations. Foreign saving, which acts as an external liquidity constraint, boosts private consumption, as shown by its significantly negative influence on saving.

Because of the lack of variation in the demographic variables over time and because the fixed-effect estimation technique does consider cross-country variation only in its country dummies, one should not attach much significance to the coefficients on the demographic variables. The urbanization rate has no discernible effect on saving under fixed-effect estimation, probably because of the collinearity between this variable and the country dummies. With OLS and random-effect estimation, the urbanization rate exerts a small negative and barely significant effect on household saving, which indicates a weak difference between urban and rural behavior that is consistent with higher consumption opportunities in cities. The dependency ratio has widely varying effects depending on the estimation technique. An interaction term between income growth and the dependency ratio was included, in the spirit of Mason and others (1986) and Collins (1989), but was found to be insignificant.

III. Conclusions and Policy Implications

The surprisingly strong results of this study, considering the small number of countries sampled, verify the value of using household data. These results should, of course, be reevaluated with a larger sample when more data become available. The empirical findings of this study confirm the central role of income and wealth in determining household saving in developing countries. Households save a larger share of their income when that income is higher and when it is growing faster. They save less the greater is their monetary wealth (compared with other forms of wealth). Borrowing constraints are also major determinants of household saving: in addition to domestic liquidity constraints (affected by current income and monetary wealth), consumers seem to face foreign liquidity constraints on the use of foreign saving. Real interest rates do not encourage saving in the countries investigated here. This may be because the substitution and income effects of higher real interest rates offset each other or because intertemporal consumption decisions are limited by liquidity constraints. As is often expected, the effect of inflation on saving is negative. This effect is at best

marginally statistically significant, however, which may result from the lack of high-inflation countries in the sample.

Table 2 summarizes the empirical results by calculating the changes in significant variables required to raise the household saving rate by one percentage point, based on the fixed-effect instrumental variables regression (equation 4 in table 1). Raising the household saving rate 1 percentage point, for example, would require an increase of 2.0 percentage points in the trend growth rate of per capita disposable income. Because the business cycle has a relatively small influence on the saving rate, a high (3.7 percentage point) increase in the ratio of current to trend disposable income is required to achieve a 1 percentage point rise in the saving rate.

Public policies can affect national saving either directly, by aiming fiscal policies at changing public sector saving, or indirectly, by inducing changes in variables that affect private saving. Although other studies have focused on the effectiveness of the first types of policies (see Corbo and Schmidt-Hebbel 1991), the results of this study may be used to derive implications for the effectiveness of indirect policies in raising household saving.

Successful stabilization policies reduce public sector deficits and hence domestic financing requirements of deficits, which lowers domestic real interest rates and inflation. Although the change in real interest rates itself has no significant effect on household saving, price stabilization could raise saving, possibly due to the reversion of past flight into consumer durables when inflation was high. A cyclical downturn caused by the stabilization measures, however, reduces household saving as long as the temporary income decline lasts.

Financial liberalization lifts interest rate controls, relaxes credit constraints, and deregulates domestic financial intermediation. Frequently this results in increased domestic real interest rates and increased monetary or financial wealth. Although the former has no discernible effect on household saving, financial deepening could depress saving as monetary wealth is raised. As shown in table 2, a 5.6 percentage point rise in the ratio of money to income would depress the saving rate by 1 percentage point. If the buildup of monetary wealth comes mostly from a portfolio substitution away from consumer durables, however, household saving could rise. It may be concluded that financial liberalization has no clear effect on household saving.

Table 2. *Changes in Saving Determinants Required to Raise Saving Rate by 1 Percentage Point*

Percentage change in the level of trend per capita disposable income	4.0
Percentage point change in the trend growth rate of per capita disposable income	2.0
Percentage point change in the ratio of current per capita disposable income to the trend level	3.7
Percentage point change in the ratio of household transfers to private disposable income	−2.4
Percentage point change in the ratio of monetary assets to private disposable income	−5.6
Percentage point change in the ratio of foreign saving to private disposable income	−8.3

Source: Authors' calculations.

Foreign capital inflows tend to reduce household saving in the short run; a fraction of each additional dollar of foreign borrowing finances private consumption. A rise in the ratio of foreign saving to income of 8.3 percentage points will result in a fall in household saving of 1 percentage point. The long-run effect of increased foreign lending on household saving is very likely positive, however, once the income growth from the investment financed by foreign lending is realized.

Successful structural adjustment and growth policies are the most effective ways to raise household saving. Reforming the incentive structure in developing countries to spur growth will feed back into higher household and, hence, national saving, thus allowing for even more economic growth. Structural reforms that move an economy beyond its initial conditions of low resource mobilization and income growth give rise to a virtuous cycle of mutually reinforcing saving and growth efforts.

References

The word "processed" describes informally reproduced works that may not be commonly available through libraries.

Arrau, Patricio. 1989. "Intertemporal Monetary Economics: Evidence from the Southern Cone of Latin America." Ph.D. dissertation. University of Pennsylvania, Department of Economics, Philadelphia. Processed.

Balassa, Bela. 1990. "The Effects of Interest Rates on Saving in Developing Countries." *Banca Nazionale del Lavoro Quarterly Review* 172 (March): 101–18.

Barro, Robert J. 1974. "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82 (6): 1095–117.

Behrman, Jere, and Chalongphob Sussangkarn. 1989. "Do the Wealthy Save Less?" Department of Economics, University of Pennsylvania. Processed.

Bernheim, B. Douglas. 1987. "Ricardian Equivalence: An Evaluation of Theory and Evidence." NBER Macroeconomics Annual 1987. Cambridge, Mass.: MIT Press.

Campbell, John, and Angus Deaton. 1989. "Why Is Consumption So Smooth?" *Review of Economic Studies* 56 (3): 357–74.

Chenery, Hollis B., and Alan M. Strout. 1966. "Foreign Assistance and Economic Development." *The American Economic Review* 56 (4, September): 679–733.

Collins, Susan M. 1989. "Savings Behavior in Ten Developing Countries." NBER Conference on Savings, Maui. National Bureau of Economic Research, Cambridge, Mass. Processed.

Corbo, Vittorio, and Klaus Schmidt-Hebbel. 1991. "Public Policies and Saving in Developing Countries." *Journal of Development Economics* 36 (1): 89–115.

Deaton, Angus. 1989. "Saving in Developing Countries: Theory and Review." *Proceedings of the World Bank Annual Conference on Development Economics, 1989*. Washington, D.C.: World Bank.

Fry, Maxwell J. 1978. "Money and Capital or Financial Deepening in Economic Development?" *Journal of Money, Credit, and Banking* 10 (4, November): 464–75.

———. 1980. "Saving, Investment, Growth, and the Cost of Financial Repression." *World Development* 8 (4): 317–27.

———. 1988. *Money, Interest, and Banking in Economic Development.* Baltimore, Md.: The Johns Hopkins University Press.

Gersovitz, Mark. 1988. "Saving and Development." In Hollis Chenery and T. N. Srinivasan, eds., *Handbook of Development Economics.* Vol. I. New York: North-Holland.

Giovannini, Alberto. 1983. "The Interest Elasticity of Savings in Developing Countries: The Existing Evidence." *World Development* 11 (7): 601–07.

———. 1985. "Saving and the Real Interest Rate in LDCs." *Journal of Development Economics* 18 (August, 2-3): 197–218.

Gupta, Kanhaya L. 1987. "Aggregate Savings, Financial Intermediation, and Interest Rate." *Review of Economics and Statistics* 69 (2): 303–11.

Hammer, Jeffrey S. 1986. "Population Growth and Savings in LDCs: A Survey Article." *World Development* 14 (5, May): 579–93.

Haque, Nadeem U., and Peter Montiel. 1989. "Consumption in Developing Countries: Tests for Liquidity Constraints and Finite Horizons." *Review of Economics and Statistics* 71 (3): 408–15.

Hayashi, Fumio. 1985. "Tests for Liquidity Constraints: A Critical Survey." NBER Working Paper 1720. National Bureau of Economic Research, Cambridge, Mass. Processed.

Hsiao, Cheng. 1986. *Analysis of Panel Data.* Cambridge: Cambridge University Press.

Hubbard, R. Glenn, and Kenneth L. Judd. 1986. "Liquidity Constraints, Fiscal Policy, and Consumption." *Brookings Papers on Economic Activity* 1: 1-50.

IMF (International Monetary Fund). Various issues. *International Financial Statistics.* Washington. D.C.

———. Various issues. *Government Financial Statistics.* Washington, D.C.

Kmenta, Jan. 1986. *Elements of Econometrics.* New York: Macmillan.

Lahiri, Ashok K. 1988. *Dynamics of Asian Savings: The Role of Growth and Age Structure.* IMF Working Paper 88/49. Washington, D.C.

Leff, Nathaniel H. 1969. "Dependency Rates and Savings Rates." *American Economic Review* 59 (5): 886–96.

Leiderman, Leonardo, and Mario I. Blejer. 1988. "Modeling and Testing Ricardian Equivalence: A Survey." *IMF Staff Papers* 35 (1): 1–35.

Mason, Andrew. 1987. "National Saving Rates and Population Growth: A New Model and New Evidence." In D. Gale Johnson and Ronald D. Lee, eds., *Population Growth and Economic Development: Issues and Evidence.* Madison: University of Wisconsin Press.

———. 1988. "Saving, Economic Growth, and Demographic Change." *Population and Development Review* 14 (1): 113–43.

Mason, Andrew, and others. 1986. *Population Growth and Economic Development: Lessons from Selected Asian Countries.* Policy Development Studies 10. New York: United Nations Fund for Population Activities.

McKinnon, Ronald I. 1973. *Money and Capital in Economic Development.* Washington D.C.: Brookings Institution.

Mikesell, Raymond F., and James E. Zinser. 1973. "The Nature of the Savings Function in Developing Countries: A Survey of the Theoretical and Empirical Literature." *Journal of Economic Literature* 11 (1): 1–26.

Rossi, Nicola. 1988. "Government Spending, the Real Interest Rate, and the Behavior of Liquidity-Constrained Consumers in Developing Countries." *IMF Staff Papers* 35 (1): 104–40.

———. 1989. "Dependency Rates and Private Savings Behavior in Developing Countries." *IMF Staff Papers* 36 (1): 166–81.

Schmidt-Hebbel, Klaus. 1987. "Foreign Shocks and Macroeconomic Adjustment in Small Open Economies." Ph.D. dissertation. Massachusetts Institute of Technology, Department of Economics, Cambridge. Processed.

Shaw, Edward S. 1973. *Financial Deepening in Economic Development*. New York: Oxford University Press.

Singh, Sateesh K. 1975. *Development Economics: Theory and Findings*. Lexington, Mass.: Lexington Books.

Taiwan, Government of. Various years. *Statistical Yearbook of the Republic of China*. Taipei.

Webb, Steven B., and Heidi Zia. 1990. "Lower Birth Rates=Higher Saving in LDCS." *Finance and Development* 27 (2): 12–14.

Williamson, Jeffrey G. 1968. "Personal Savings in Developing Nations: An Intertemporal Cross-Section from Asia." *Economic Record* 44 (106, June): 194–210.

Zeldes, Stephen P. 1989. "Consumption and Liquidity Constraints: An Empirical Investigation." *Journal of Political Economy* 97 (2): 305–46.

Cumulative Index of Titles

Volume 6 (1992)

THE WORLD BANK ECONOMIC REVIEW, VOL 6, NO 3: 551-552

Cumulative Index of Authors

VOLUME 6 (1992)

Coming in the next issue of

THE WORLD BANK ECONOMIC REVIEW

January 1993
Volume 7, Number 1

Articles on . . .

- **External Shocks, Purchasing Power Parity, and the Equilibrium Real Exchange Rate**
  by Shantayanan Devarajan, Jeffrey D. Lewis, and Sherman Robinson

- **Moderate Inflation**
  by Rudiger Dornbusch and Stanley Fischer

- **Theoretical Implications of Imperfect Competition on Quota License Prices and Auctions**
  by Kala Krishna

- **Obstacles to the Development of Indigenous Small and Medium Enterprises**
  by Brian Levy

- **Estimating Returns to Scale with Large Imperfect Panels: An Application to Chilean Manufacturing Industries**
  by James R. Tybout and M. Daniel Westbrook